Instrumental variable estimation via a continuum of instruments with an application to estimating the elasticity of intertemporal substitution in consumption

Xuexin WANG

April 5, 2023

Abstract

This study proposes new instrumental variable (IV) estimators for linear models, effectively utilizing a continuum of instruments. The effectiveness is attributed to the unique weighting function employed in the minimum distance objective functions. The proposed estimators are robust to weak instruments and heteroscedasticity of unknown form. Moreover, they are robust to high dimensionality of included and excluded exogenous variables. The proposed estimators have analytical formulas, which are easily computable. Inference drawn from these estimators is also straightforward, as their variance estimators for parameter inferences are also analytical. Comprehensive Monte Carlo simulations confirm that the proposed estimators exhibit excellent finite sample properties and outperform alternative estimators over a wide range of cases. The new estimation procedure is applied for estimating the elasticity of intertemporal substitution (EIS) in consumption, which is of central importance in macroeconomics and finance. For quarterly data of the US from Q4 1955 to Q1 2018, the EIS estimates obtained through our approach exceed one and are statistically significant. These findings persist across model transformations, different sets of IVs, data structures, and data ranges.

JEL Classification:. C12, C13, C23.

Keywords: Endogeneity; Heteroskedasticity of unknown form; Jackknife; Weak identification; EIS in consumption

Address: The Wang Yanan Institute for Studies in Economics(WISE), Xiamen University, Xiamen, Fujian, China. E-mail: *xuexinwang@xmu.edu.cn*.

1 Introduction

In econometrics, there is a significant body of literature on instrumental variable (IV) methods that aim to address the endogeneity problem in linear models. Generally, IVs must satisfy the exogeneity and relevance conditions, yet when the strength of instrumental relevance is weak, we can encounter the issue of weak instruments or weak identification, which is quite pervasive in economic applications.

Some IV estimators, like two-stage least squares (TSLS), are known to be susceptible to weak instruments. The implications of weak instruments on TSLS are highlighted in Nelson and Startz (1990) and Bound et al. (1995). According to Staiger and Stock (1997), the TSLS and limited information maximum likelihood (LIML) estimators are theoretically inconsistent and converge instead to non-standard distributions in a $n^{-1/2}$ local-to-zero parametrization in the first-stage regression, where n represents the sample size.

While the assumption that the number of instruments is fixed underpins the conclusions of Staiger and Stock (1997), Chao and Swanson (2005) revealed that increasing the number of instruments can improve the estimation accuracy of the LIML and bias-corrected two-stage least square (BTSLS) estimators in the presence of weak instruments, although TSLS remains inconsistent. Hansen et al. (2008) formulated corrected standard errors for the LIML estimator and the Fuller (1977) (FULL) estimator in such cases. However, as pointed out by Bekker and van der Ploeg (2005) and Hausman et al. (2012), the asymptotic consistency of LIML and FULL, under many weak instruments asymptotics, breaks down in the presence of heteroskedasticity of unknown form. To address this issue, Hausman et al. (2012) propose the heteroskedasticity-robust version of the FULL (HFUL) estimator, which is based on a jackknife version of the LIML estimator, referred to as HLIM. They demonstrate HFUL outperforms alternative estimators, such as the jackknife IV estimators (JIVE) developed by Phillips and Hale (1977), Blomquist and Dahlberg (1999), Angrist et al. (1999), and Ackerberg and Devereux (2009). It is worth mentioning that the existing studies on many weak instruments originate from a large body of literature on many instruments, such as Morimune (1983) and Bekker (1994). See also the comprehensive survey of Anatolyev (2019).

However, determining an appropriate number of instruments for the standard many weak IV estimators is extremely challenging in practice. In fact, even in the presence of strong instruments, it is a delicate task to select the correct number of instruments in linear models, as highlighted in

Morimune (1983), Donald and Newey (2001), and Carrasco and Tchuente (2015), among others. When a linear reduced form in the first regression is assumed, the asymptotic properties of the standard many weak IV estimators crucially depend on the interplay between the number of instruments and their strength, as demonstrated by Chao and Swanson (2005). Unfortunately, the reduced form is totally unknown in most cases. The standard many weak IV estimators further require that the linear combination of an increasing number of instruments should approximate the reduced form sufficiently as the sample size goes to infinity. However, in the presence of weak instruments, incorporating more instruments can lead to more accurate IV estimates, but including too many can result in an increase in the bias and variance of the reduced form estimator in finite samples, hence deteriorating the accuracy of the estimates.

In this study, by utilizing a full continuum of instruments effectively, we propose new nuisance parameter-free IV estimators. Therefore, they conveniently address the limitations of the standard many weak IV estimators. Remarkably, the proposed estimators maintain analytical formulas and have a natural jackknife form, resembling HLIM and HFUL, respectively. We label the HLIM-like estimator as WCIV, as its objective function involves a weighted continuum of IVs, and label the Fuller-like version of WCIV as WCIVF. We demonstrate that WCIV and WCIVF are consistent and asymptotically normally distributed in the presence of weak instruments and heteroskedasticity of unknown form. The inference drawn from these estimators is also straightforward. Comprehensive Monte Carlo simulations reveal that WCIV and WCIVF outperform HFUL and other competitors in a wide range of cases. We use WCIV and WCIVF to estimate the elasticity of intertemporal substitution (EIS) in consumption based on macro datasets from the US. For the quarterly data ranging from Q4 1955 to Q1 2018, the WCIV and WCIVF estimates of EIS are well above one and statistically different from zero. These findings are robust to model transformation, different sets of IVs, different data structures and data ranges.

This study makes two main contributions. Firstly, it provides an elegant solution for estimating linear models with weak instruments and heteroskedasticity of unknown form, which are defined in terms of conditional moment restrictions. In this scenario, choosing an appropriate number of moments for standard many weak IV estimators is extremely challenging. The uniqueness of our approach lies in employing a novel non-integrable weighting function in the minimum distance objective functions formulated from the continuum of IVs. This weighting function enjoys some attractive features. One outstanding feature is that its weighting values in a neighborhood of the origin tend to be infinite. This is extremely important in terms of estimation efficiency, as the sample moments generated from the continuum of IVs are most informative in this neighborhood. Moreover, this weighting function is an increasing function of the dimension of included and excluded exogenous variables, making estimators robust to high-dimensionality. This feature is also important because in the presence of weak instruments, it is advantageous to include more excluded exogenous variables to augment the instrumental variable relevance, and include more included exogenous variables to safeguard against model misspecification, or to approximate unobservable factors. In addition, through this weighting function, the minimum distance objective functions and, consequently, WCIV and WCIVF enjoy analytical forms; therefore, they are easily computable. Lastly, under this weighting function, the objective functions are of jackknife representation, which ensures that WCIV and WCIVF are robust to heteroskedasticity of unknown form. To the best of our knowledge, no previous weighting function has demonstrated all the above properties simultaneously.

Secondly, the estimates of the EIS in consumption obtained by WCIV and WCIVF suggest a resolution to a long-standing discrepancy between EIS values in many model calibrations and those estimated via macro datasets. Previous empirical studies, such as Hall (1988), Campbell (2003), Yogo (2004), and Ascari et al. (2021), have obtained small EIS values. On the other hand, the EIS in consumption in many model calibrations is required to be significantly large to accord with the stylized facts of macroeconomic dynamics.

A continuum of instruments (moments) has been utilized in consistent specification tests for models defined by conditional moment restrictions; see Bierens (1990) and Bierens and Ploberger (1997), among others. Similarly, a continuum of moments has been utilized in estimation procedures for models defined by conditional moment restrictions, see, Domínguez and Lobato (2004) and Hsu and Kuan (2011). These studies mainly focus on the consistent parameter estimation of nonlinear models under minimal global identifying conditions. For linear models, Escanciano (2018) and Antoine and Lavergne (2014) utilize a continuum of moments that is similar to the one in this study. Their minimum distance objective functions involve integrable weighing functions. Their IV estimators are generally inferior to or comparable to HFUL as observed in Antoine and Lavergne (2014), and worse than WCIV and WCIVF, as demonstrated in this study. On the other hand, Carrasco and Florens (2000) establish an estimation framework involving a continuum of moments, extending the generalized method of moments (GMM) of Hansen (1982).

In order to pursue estimation efficiency, their minimum distance objective function involves a random weighting function (covariance operator), which is analogous to the optimal weighting matrix in GMM. This approach depends on a regularization of the covariance operator to solve an ill-posed problem.

The remainder of the paper is organized as follows. Section 2 introduces the model setup and new IV estimators. We provide the simple analytical formulas for WCIV and WCIVF, the variance estimators and, the valid Wald test statistic. Section 3 introduces the nonintegrable weighting function, and the minimum distance objective functions. Section 4 establishes the asymptotic theory of our proposed IV estimators. Section 5 conducts a comprehensive Monte Carlo simulation study. Section 6 presents the application of estimating the EIS in consumption. Section 7 concludes. The proofs are presented in the Appendix.

Throughout the paper, for a complex-valued function $f(\cdot)$, its complex conjugate is denoted by $f^c(\cdot)$ and $|f(\cdot)|^2 = f(\cdot) f^c(\cdot)$. The scalar product of vectors $\boldsymbol{\tau}$ and $\boldsymbol{\varsigma}$ in a Euclidean space is denoted by $\langle \boldsymbol{\tau}, \boldsymbol{\varsigma} \rangle$. The Euclidean norm of $\mathbf{X} = (X_1, ..., X_q)$ in $\mathbb{C}^q$ is $\|\mathbf{X}\|$, where $\|\mathbf{X}\|^2 = \sum_{j=1}^q X_j X_j^c$. Variables $\mathbf{X}^+$ and $\mathbf{X}^{++}$ are independent copies of $\mathbf{X}$, that is, $\mathbf{X}^+$, $\mathbf{X}^{++}$ and $\mathbf{X}$ are independent and identically distributed (i.i.d.). For a matrix $\mathbf{X}$, $\mathbf{X}'$ is its transpose matrix. Let $\vartheta_{\min}(\mathbf{A})$ denote the smallest eigenvalue of a symmetric matrix $\mathbf{A}$.

2 Model Setup and New IV Estimators

Consider the following model

$$y_t = \alpha_0 + \boldsymbol{\beta}_0' \mathbf{Y}_t + \varepsilon_t,\ t = 1, ..., n,$$

where $\mathbf{Y}_t$ is a $p \times 1$ vector of regressors, which is potentially correlated with the error term ε_t; $\boldsymbol{\theta}_0 = \left(\alpha_0, \boldsymbol{\beta}_0'\right)' \subset \mathbb{R}^{1+p}$. The IV regression approach assumes that there exists a $q \times 1$ dimensional vector of exogenous variables $\mathbf{X}_t$ (excluding a constant), $q \geq p$, such that, almost surely (a.s.)

$$E\left(\varepsilon_t | \mathbf{X}_t\right) = 0. \tag{1}$$

In this setup, $\mathbf{Y}_t$ contains the included exogenous variables. Correspondingly, $\mathbf{X}_t$ contains these variables in addition to the excluded exogenous variables. Condition (1) for instrumental exogeneity is a conditional moment restriction that appears regularly in macroeconomic and financial econometric models, such as log-linearized Euler equations of asset pricing models, dynamic panel data models, and new Keynesian Phillips curves, to name a few.

With the exogeneity condition being satisfied, the formal identification of the parameter $\boldsymbol{\beta}_0$ depends on the conditional expectation $E(\mathbf{Y}_t|\mathbf{X}_t)$. In this context, consider two distinct parameters $(\alpha_1, \boldsymbol{\beta}_1')'$ and $(\alpha_2, \boldsymbol{\beta}_2')'$; they are observationally equivalent if and only if

$$E\left(y_t - \alpha_1 - \boldsymbol{\beta}_1'\mathbf{Y}_t|\mathbf{X}_t\right) = E\left(y_t - \alpha_2 - \boldsymbol{\beta}_2'\mathbf{Y}_t|\mathbf{X}_t\right),$$

or

$$(\alpha_1 - \alpha_2) + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)' E(\mathbf{Y}_t|\mathbf{X}_t) = 0.$$

Clearly the identification strength of $\boldsymbol{\beta}_0$ directly depends on $E[\mathbf{Y}_t|\mathbf{X}_t]$, while α_0 is always strongly identified. When $E[\mathbf{Y}_t|\mathbf{X}_t]$ flattens to zero as the sample size increases (cf. Assumption 2 in Section 4), the IV estimate of $\boldsymbol{\beta}_0$ may suffer from the weak identification problem.

This study utilizes a continuum of instruments, such that

$$\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right), \text{for all } \boldsymbol{\tau} \in \mathbb{R}^q.$$

Precisely, we employ the following continuum of unconditional moment restrictions:

$$E\left\{\left[y_t - \mu_y - \boldsymbol{\beta}_0'(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\right]\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right\} = 0, \text{for all } \boldsymbol{\tau} \in \mathbb{R}^q, \tag{2}$$

where $\mu_y = E(y_t)$ and $\boldsymbol{\mu}_Y = E(\mathbf{Y}_t)$. It is observed that α_0 is canceled out. In this study the focus is on $\boldsymbol{\beta}_0$. Clearly, there exists an equivalence between (1) and (2).

Although we utilize a continuum of instruments, our proposed estimators WCIV and WCIVF enjoy convenient analytical formulas. To describe them, let $\mathbf{Y} = [\mathbf{Y}_1, ..., \mathbf{Y}_n]'$, $\mathbf{y} = [y_1, ..., y_n]'$, $\bar{\mathbf{Y}} = \frac{1}{n}\sum_{t=1}^n \mathbf{Y}_t$, $\bar{y} = \frac{1}{n}\sum_{t=1}^n y_t$. Define

$$\tilde{\mathbf{Y}} = \left[\mathbf{Y}_1 - \bar{\mathbf{Y}}, ..., \mathbf{Y}_n - \bar{\mathbf{Y}}\right]'$$

and

$$\tilde{\mathbf{y}} = [y_1 - \bar{y}, ..., y_n - \bar{y}]'.$$

Let $\mathbf{D}$ be a square matrix of size n and D_{jk} denote the (j,k)th element of $\mathbf{D}$, such that

$$D_{jk} = -\left\|\mathbf{X}_j - \mathbf{X}_k\right\|, \, j,k = 1,...,n.$$

The WCIV estimator is given as

$$\hat{\boldsymbol{\beta}}_{WCIV} = \left[\tilde{\mathbf{Y}}'\left(\mathbf{D} - \hat{\lambda}_{WCIV}\mathbf{I}_n\right)\tilde{\mathbf{Y}}\right]^{-1}\left[\tilde{\mathbf{Y}}'\left(\mathbf{D} - \hat{\lambda}_{WCIV}\mathbf{I}_n\right)\tilde{\mathbf{y}}\right] \tag{3}$$

$$\hat{\alpha}_{WCIV} = \bar{y} - \hat{\boldsymbol{\beta}}'_{WCIV}\bar{\mathbf{Y}}, \tag{4}$$

where $\mathbf{I}_n$ is an identity matrix of size n and

$$\hat{\lambda}_{WCIV} \text{ is the smallest eigenvalue of } \left(\check{\mathbf{Y}}'\check{\mathbf{Y}}\right)^{-1}\check{\mathbf{Y}}\mathbf{D}\check{\mathbf{Y}}$$

with $\check{\mathbf{Y}} = \left[\tilde{\mathbf{y}}, \tilde{\mathbf{Y}}\right]$. The WCIVF estimator is written as (3), replacing $\hat{\lambda}_{WCIV}$ with

$$\left[\hat{\lambda}_{WCIV} - \left(1 - \hat{\lambda}_{WCIV}\right)C/n\right] / \left[1 - \left(1 - \hat{\lambda}_{WCIV}\right)C/n\right],$$

where C is a constant.

Clearly, WCIV and WCIVF resemble HLIM and HFUL, respectively, as $D_{jj} = 0$ for $j = 1,..,n$. Recall that conventional k-class IV estimators are of the form

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \left[\mathbf{Y}^{*\prime}\left(\mathbf{P} - \hat{\lambda}\mathbf{I}_n\right)\mathbf{Y}^*\right]^{-1}\left[\mathbf{Y}^{*\prime}\left(\mathbf{P} - \hat{\lambda}\mathbf{I}_n\right)\mathbf{y}\right],$$

where $\mathbf{Y}^* = [\boldsymbol{\iota}, \mathbf{Y}]$, $\boldsymbol{\iota}$ is the vector of ones, and $\mathbf{P}$ is a matrix that depends on $n \times m$ matrix $\mathbf{Z}$ of instrumental variable observations with $rank\,(\mathbf{Z}) = m \geq p + 1$. TSLS corresponds to $\mathbf{P} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$, and $\hat{\lambda} = 0$; JIVE corresponds to $\mathbf{P} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}' - diag\left(\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right)$ and $\hat{\lambda} = 0$; LIML corresponds to $\mathbf{P} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$, and $\hat{\lambda}$ equals to the smallest eigenvalue of $\left(\check{\mathbf{Y}}^{*\prime}\check{\mathbf{Y}}^*\right)^{-1}\check{\mathbf{Y}}^*\mathbf{P}\check{\mathbf{Y}}^*$ with $\check{\mathbf{Y}}^* = [\mathbf{y}, \mathbf{Y}^*]$; HLIM corresponds to $\mathbf{P} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}' - diag\left(\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right)$, $\hat{\lambda}$ equals to the smallest eigenvalue of $\left(\check{\mathbf{Y}}^{*\prime}\check{\mathbf{Y}}^*\right)^{-1}\check{\mathbf{Y}}^{*\prime}\mathbf{P}\check{\mathbf{Y}}^*$. Finally, HFUL employs

$$\hat{\lambda}_{HFUL} = \left[\hat{\lambda}_{HLIM} - \left(1 - \hat{\lambda}_{HLIM}\right)C/n\right] / \left[1 - \left(1 - \hat{\lambda}_{HLIM}\right)C/n\right]$$

in HLIM.

Moreover, the valid Wald test statistic for parameter inference is easily computable. Consider testing the parametric restriction of the form

$$H_0 : \mathbf{g}(\boldsymbol{\beta}_0) = 0, \tag{5}$$

where $\mathbf{g}(\cdot)$ is a function from $\mathbb{R}^p$ on $\mathbb{R}^m$ with $m \leq p$.

To describe the Wald statistic, let $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}')'$, $\varepsilon_t(\boldsymbol{\theta}) = y_t - \alpha - \boldsymbol{\beta}'\mathbf{Y}_t$, $\tilde{\mathbf{Y}}_t = \mathbf{Y}_t - \bar{\mathbf{Y}}$, $\tilde{\mathbf{D}}(\lambda) = \mathbf{D} - \lambda \mathbf{I}_n$ and $\tilde{D}_{jk}(\lambda)$ denote the (j,k)th element of $\tilde{\mathbf{D}}(\lambda)$. Define

$$\hat{\mathbf{S}}_1(\boldsymbol{\theta}, \lambda) = \frac{1}{n^3} \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \tilde{\mathbf{Y}} j \tilde{\mathbf{Y}}'_k \varepsilon_l(\theta)^2 \tilde{D}_{jl}(\lambda) \tilde{D}_{kl}(\lambda),$$

$$\hat{\mathbf{S}}_2(\boldsymbol{\theta}, \lambda) = \frac{1}{n^5} \sum_{l=1}^n \varepsilon_l(\theta)^2 \left(\sum_{j=1}^n \sum_{k=1}^n \tilde{\mathbf{Y}}_j \tilde{D}_{jk}(\lambda) \right) \left(\sum_{j=1}^n \sum_{k=1}^n \tilde{\mathbf{Y}}'_j \tilde{D}_{jk}(\lambda) \right),$$

$$\hat{\mathbf{S}}_3(\boldsymbol{\theta}, \lambda) = \frac{1}{n^4} \sum_{j=1}^n \sum_{k=1}^n \varepsilon_k(\theta)^2 \tilde{\mathbf{Y}}_j \tilde{D}_{jk}(\lambda) \sum_{j=1}^n \sum_{k=1}^n \tilde{\mathbf{Y}}'_j \tilde{D}_{jk}(\lambda);$$

further,

$$\hat{\boldsymbol{\Omega}}(\boldsymbol{\theta}, \lambda) = \hat{\mathbf{S}}_1(\theta, \lambda) + \hat{\mathbf{S}}_2(\theta, \lambda) - \hat{\mathbf{S}}_3(\theta, \lambda) - \hat{\mathbf{S}}'_3(\theta, \lambda),$$

and

$$\hat{\boldsymbol{\Upsilon}}(\lambda) = \frac{1}{n^2} \tilde{\mathbf{Y}}' \tilde{\mathbf{D}}(\lambda) \tilde{\mathbf{Y}}.$$

The Wald test statistic is constructed as

$$W_n\left(\hat{\boldsymbol{\theta}}\right) = n\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)' \left(\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right) \hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right) \mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)' \right)^{-1} \mathbf{g}\left(\hat{\boldsymbol{\beta}}\right), \tag{6}$$

where $\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right) = \partial \mathbf{g}\left(\hat{\boldsymbol{\beta}}\right) / \partial \boldsymbol{\beta}'$, $\hat{\mathbf{V}}(\boldsymbol{\theta}, \lambda) = \hat{\boldsymbol{\Upsilon}}(\lambda)^{-1} \hat{\boldsymbol{\Omega}}(\boldsymbol{\theta}) \hat{\boldsymbol{\Upsilon}}(\lambda)^{-1}$.

It is observed that $\hat{\mathbf{V}}(\boldsymbol{\theta}, \lambda)$ has a sandwich form and is easy to compute, unlike Hausman et al. (2012) where an extra term is included to account for the numerosity of instruments. It is worth mentioning that $\hat{\mathbf{V}}(\boldsymbol{\theta}, \lambda)/n$ is not a consistent variance estimator for the population variance of $\hat{\boldsymbol{\beta}}_{WCIV}$ or $\hat{\boldsymbol{\beta}}_{WCIVF}$ under weak instruments, which involves the unknown degrees of weak identification.

3 Nonintegrable Weighting Function and Objective Functions

To fully utilize the continuum of moments (2) we have introduced in previous section, a distance measure is to be formulated. Denote

$$h(\boldsymbol{\beta}, \boldsymbol{\tau}) = E\left\{\left[y_t - \mu_y - \boldsymbol{\beta}'(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\right] \exp\left(i \langle \boldsymbol{\tau}, \mathbf{X}_t \rangle\right)\right\},$$

and its sample analog

$$h_n(\boldsymbol{\beta}, \boldsymbol{\tau}) = \frac{1}{n} \sum_{t=1}^{n} \left(\tilde{y}_t - \boldsymbol{\beta}'\tilde{\mathbf{Y}}_t\right) \exp\left(i \langle \boldsymbol{\tau}, \mathbf{X}_t \rangle\right),$$

where $\tilde{y}_t = y_t - \bar{y}$. The distance measure has the form

$$\int_{\mathbb{R}^q} |h(\boldsymbol{\beta}, \boldsymbol{\tau})|^2 W(d\boldsymbol{\tau})$$

and its sample analog is

$$\int_{\mathbb{R}^q} |h_n(\boldsymbol{\beta}, \boldsymbol{\tau})|^2 W(d\boldsymbol{\tau}),$$

where $W(\cdot)$ is an arbitrary positive weighting function for which the integrals mentioned above exist.

Clearly, $W(\cdot)$ is pivotal in terms of estimation accuracy, as it acts similar to the weighting matrix in the objective function of GMM in Hansen (1982). Different choices of $W(\cdot)$ give rise to associated IV estimators with different asymptotic properties. It is possible to introduce a random weighting function, following Carrasco and Florens (2000). However, under weak instruments and heteroskedasticity of unknown form, it is extremely challenging to follow this approach. Further, their approach requires a regularization of the weighting function, which involves a tuning parameter and is quite difficult to implement in practice. Instead, our approach employs a nonrandom weighting function in the distance measure, which avoids the challenging issue regarding the selection of an appropriate number of instruments or a tuning parameter. One main contribution of this study is to introduce a unique nonintegrable weighting function, which renders the proposed estimators outstanding theoretical and empirical properties, deviating substantially from Escanciano (2018) and Antoine and Lavergne (2014), where some integrable weighting functions, such as the standard normal density function, are employed.

Intuitively, from the perspective of the estimation efficiency of GMM, more weighting values

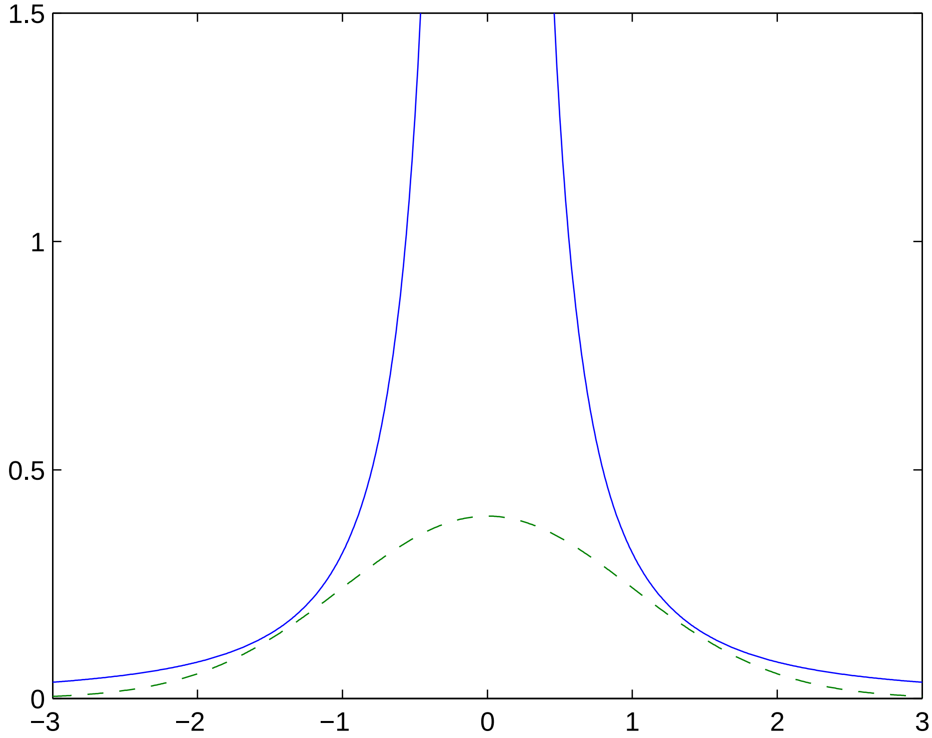

Figure 1: Standard normal density function (dashed curve) vs. $1/(\pi\boldsymbol{\tau}^2)$ (solid curve)

should be attributed to more informative sample moments in the minimum distance objective function. When $\boldsymbol{\tau}$ is in a neighborhood of the origin, $h_n(\boldsymbol{\beta}, \boldsymbol{\tau})$ contains more information, as it can be shown, under some regularity conditions,

$$E\left|\sqrt{n}h_n(\boldsymbol{\beta}_0, \boldsymbol{\tau})\right|^2 = \frac{2(n-1)^2}{n^2} E\left\{\varepsilon_t^2\left[1 - \cos\left(\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+ \rangle\right)\right]\right\}.$$

Hence, weighting values as high as possible in a neighborhood of the origin is preferred. To this end, we employ a nonintegrable weighting function, such that

$$W(\boldsymbol{\tau}) = \frac{1}{c_q \|\boldsymbol{\tau}\|^{q+1}}, \tag{7}$$

where c_q is a constant defined in Lemma 8.1 in the Appendix. One outstanding feature of (7) is that its weighting values tend to infinity as $\|\boldsymbol{\tau}\| \to 0$, which is strikingly different from integrable weighting functions, such as a standard normal density function. Figure 1 demonstrates this fact for a standard normal density function, and (7) with $q = 1$.

Another important feature of (7) is that it is an increasing function of q. Note that

$$
\begin{aligned}
\left(c_q \left\|\boldsymbol{\tau}\right\|^{q+1}\right)^{-1} &= \frac{\Gamma\left(\left(q+1\right)/2\right)}{\pi^{(q+1)/2}} \left(\left\|\boldsymbol{\tau}\right\|^{q+1}\right)^{-1} \\
&= \frac{(q-1)!!}{(2\pi)^{q/2} \left\|\boldsymbol{\tau}\right\|^{q+1}},
\end{aligned}
$$

where $q!!$ is a double factorial, such that

$$
q!! = \begin{cases} q \cdot (q-2)\ldots 5 \cdot 3 \cdot 1 & q > 0 \text{ odd} \\ q \cdot (q-2)\ldots 6 \cdot 4 \cdot 2 & q > 0 \text{ even} \\ 1 & q = -1, 0. \end{cases}
$$

By applying an approximation to $(q-1)!!$, when $q > 1$, we have

$$
\begin{aligned}
\left(c_q \left\|\boldsymbol{\tau}\right\|^{q+1}\right)^{-1} &\approx c\left(q-1\right)^{q/2} e^{-(q-1)/2} \frac{1}{(2\pi)^{q/2} \left\|\boldsymbol{\tau}\right\|^{q+1}} \\
&\approx c\sqrt{e}\left(\frac{q-1}{2\pi e}\right)^{q/2} \left(\left\|\boldsymbol{\tau}\right\|^{q+1}\right)^{-1} \\
&\approx \frac{c\sqrt{e}}{\left\|\boldsymbol{\tau}\right\|}\left(\frac{q-1}{2\pi e \left\|\boldsymbol{\tau}\right\|^2}\right)^{q/2},
\end{aligned}
$$

where $c = \sqrt{\pi}$ for $q-1$ is even and $\sqrt{2}$ for $q-1$ is odd. Therefore, for a fixed value $\|\boldsymbol{\tau}\|$ in a neighborhood of the origin, (7) is an increasing function of q. This is important, as in weak instruments scenarios, it is well motivated to introduce more excluded exogenous variables to improve the IV strength in addition to the fact that many exogenous regressors are typically included to guard against model misspecification or approximate some important but unobservable factors. On the other hand, a q-dimensional standard normal density function is a decreasing function of q, given a $\|\boldsymbol{\tau}\|$. Notably, its weighting value equals $(2\pi)^{-q/2}$ at the origin, being the maximum. It sharply shrinks to zero when q increases.

The nonintegrable weighting function was first introduced by Székely et al. (2007) in the statistics literature. Studies involving this weighting function include Székely and Rizzo (2009), Székely and Rizzo (2014), Shao and Zhang (2014), Davis et al. (2018), Zhang et al. (2018), Yao et al. (2018), and Wang (2021) in testing framework. In the following, write

$$
\int_{\mathbb{R}^q} \frac{\left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2}{c_q \left\|\boldsymbol{\tau}\right\|^{q+1}} d\boldsymbol{\tau} = \int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right),
$$

where $\boldsymbol{\omega}(d\boldsymbol{\tau}) = \left(c_q \|\boldsymbol{\tau}\|^{q+1}\right)^{-1} d\boldsymbol{\tau}$ for notational simplicity. The third feature of this nonintegrable weighting function is that $\int_{\mathbb{R}^q} |h(\boldsymbol{\beta}, \boldsymbol{\tau})|^2 \boldsymbol{\omega}(d\boldsymbol{\tau})$ enjoys a convenient analytical form, as demonstrated by Lemma 3.1.

Lemma 3.1 *Under Assumptions 1-3 presented in the next section, for any* $\boldsymbol{\beta} \in \mathbb{R}^p$*,*

$$\int_{\mathbb{R}^q} |h(\boldsymbol{\beta}, \boldsymbol{\tau})|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) = -E\left[\left(y_t - \mu_y - \boldsymbol{\beta}'(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\right)\left(y_t^+ - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right], \tag{8}$$

where $\left(y_t^+, (\mathbf{X}_t^+)'\right)'$ *is an i.i.d. copy of* $(y_t, \mathbf{X}_t')'$*.*

Proof. See the Appendix. ■

It is possible to obtain a new IV estimator by minimizing the sample analog of (8). However, preliminary Monte Carlo simulations have revealed that this estimator can be very biased in the presence of weak instruments and a high dimensionality of $\mathbf{X}_t$. To improve the estimation accuracy under these circumstances, we construct a LIML-like objective function, which is a weighted version of (8), as follows:

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}} \frac{\int_{\mathbb{R}^q} |h(\boldsymbol{\beta}, \boldsymbol{\tau})|^2 \boldsymbol{\omega}(d\boldsymbol{\tau})}{E\left(\left[y_t - \mu_y - \boldsymbol{\beta}'(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\right]^2\right)}, \tag{9}$$

$$\alpha_0 = \mu_y - \boldsymbol{\beta}_0' \boldsymbol{\mu}_Y. \tag{10}$$

Then, the WCIV estimator is defined as the minimizer of the sample analog of (9), such that

$$\hat{\boldsymbol{\beta}}_{WCIV} = \arg\min_{\boldsymbol{\beta}} \left[\frac{\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)' \mathbf{D} \left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)}{\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)' \left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)}\right], \tag{11}$$

$$\hat{\alpha}_{WCIV} = \bar{y} - \hat{\boldsymbol{\beta}}_{WCIV}' \bar{\mathbf{Y}}. \tag{12}$$

Obtaining (3) is straightforward and analogous to the computation of HLIM. Moreover, WCIV remains invariant to normalization, similar to the case of HLIM. However, like HLIM, WCIV may suffer from the moments problem in some cases. To address this issue, this study suggests utilizing a Fuller-type finite sample correction of WCIV (WCIVF), following the approach presented in Fuller (1977), Hahn et al. (2004), and Hausman et al. (2012). WCIVF is obtained directly by replacing $\hat{\lambda}_{WCIV}$ in the WCIV estimator (3) with $\left[\hat{\lambda}_{WCIV} - \left(1 - \hat{\lambda}_{WCIV}\right) C/n\right] / \left[1 - \left(1 - \hat{\lambda}_{WCIV}\right) C/n\right]$.

It is worth noting that both WCIV and WCIVF possess natural jackknife representation, making them robust to heteroskedasticity of unknown form.

It is worth mentioning that Antoine and Lavergne (2014) use an integrable weighting function in objective functions, based on a continuum of moments:

$$E\left\{\left[y_t - \boldsymbol{\theta}_0' \mathbf{Y}_t^*\right] \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right\} = 0, \text{for all } \boldsymbol{\tau} \in \mathbb{R}^q,$$

where $\mathbf{Y}_t^* = (1, \mathbf{Y}_t')'$. When a standard normal density function is chosen, the MD estimator is calculated as

$$\hat{\boldsymbol{\theta}}_{MD} = \arg\min_{\theta} \left[(\mathbf{y} - \mathbf{Y}^*\boldsymbol{\theta})' \mathbf{K} (\mathbf{y} - \mathbf{Y}^*\boldsymbol{\theta})\right],$$

where $\mathbf{K}$ is a $n \times n$ matrix, such that $K_{jk} = \exp\left(-\left\|\mathbf{X}_j - \mathbf{X}_k\right\|^2 /2\right)$ for $j \neq k$, and $K_{jj} = 0$ for $j, k = 1, ..., n$. Note that $\exp\left(-\left\|\mathbf{X}_j - \mathbf{X}_k\right\|^2 /2\right) = 1 \neq 0$, when $j = k$. Therefore the diagonal elements of $\mathbf{K}$ need to be set to zero to form a jackknife form.[1] The WMD estimator is

$$\hat{\boldsymbol{\theta}}_{WMD} = \arg\min_{\boldsymbol{\theta}} \left[\frac{(\mathbf{y} - \mathbf{Y}^*\boldsymbol{\theta})' \mathbf{K} (\mathbf{y} - \mathbf{Y}^*\boldsymbol{\theta})}{(\mathbf{y} - \mathbf{Y}^*\boldsymbol{\theta})' (\mathbf{y} - \mathbf{Y}^*\boldsymbol{\theta})}\right].$$

Note that in Antoine and Lavergne (2014), the full parameter vector θ can be estimated by minimizing the objective function. Consequently, the formula of $\hat{\theta}_{MD}$ and $\hat{\theta}_{WMD}$ can be obtained, such that

$$\hat{\boldsymbol{\theta}}_{MD} = \left[\mathbf{Y}^{*\prime} K \mathbf{Y}^*\right]^{-1} \left[\mathbf{Y}^{*\prime} \mathbf{K} \mathbf{y}\right],$$

$$\hat{\boldsymbol{\theta}}_{WMD} = \left[\mathbf{Y}^{*\prime} \left(\mathbf{K} - \hat{\lambda}_{WMD} \mathbf{I}_n\right) \mathbf{Y}^*\right]^{-1} \left[\mathbf{Y}^{*\prime} \left(\mathbf{K} - \hat{\lambda}_{WMD} \mathbf{I}_n\right) \mathbf{y}\right], \tag{13}$$

where $\hat{\lambda}_{WMD}$ is the minimum value of the objective function, which can be explicitly computed as the smallest eigenvalue of $\left(\check{\mathbf{Y}}^{*\prime}\check{\mathbf{Y}}^*\right)^{-1} \check{\mathbf{Y}}^{*\prime} K \check{\mathbf{Y}}^*$ with $\check{\mathbf{Y}}^* = [\mathbf{y}, \mathbf{Y}^*]$. Its Fuller-style variant WMDF is obtained directly by replacing $\hat{\lambda}_{WMD}$ in the WMD estimator (13) with

$$\left[\hat{\lambda}_{WMD} - \left(1 - \hat{\lambda}_{WMD}\right) C/n\right] / \left[1 - \left(1 - \hat{\lambda}_{WMD}\right) C/n\right].$$

It appears that both WMD and WCIV (WMDF and WCIVF) share many similarities, but they are constructed on distinct estimation frameworks. Due to the fundamental difference between the weighting functions in the objective functions, WCIV and WCIVF are expected to

[1] The MD estimator, without setting zero values for the diagonal elements of K, corresponds to the IV estimator proposed by Escanciano (2018).

be less dispersed than WMD and WMDF in finite samples.

4 Asymptotic Theory

To appreciate the asymptotic theory of WCIV and WCIVF, we introduce some assumptions.

Assumption 1 *Let* $\mathbf{W}_t$ *denote a vector containing distinct elements of* $(y_t, \mathbf{Y}_t', \mathbf{X}_t')'$. $\{\mathbf{W}_t\}_{t=1}^n$ *are i.i.d.* $E\left\|\mathbf{X}_t\right\|^2 < \infty$.

Assumption 2 $\mathbf{Y}_t = E\left(\mathbf{Y}_t|\mathbf{X}_t\right) + \boldsymbol{\eta}_t$.

$$E\left(\mathbf{Y}_t|\mathbf{X}_t\right) = \frac{\mathbf{R}_n \mathbf{f}\left(\mathbf{X}_t\right)}{\sqrt{n}},$$

where $\mathbf{R}_n = \tilde{\mathbf{R}}_n diag\left(r_{1,n}, ..., r_{q,n}\right)$, *such that* $\tilde{\mathbf{R}}_n$ *is bounded, and the smallest eigenvalue of* $\tilde{\mathbf{R}}_n\tilde{\mathbf{R}}_n'$ *is bounded away from zero. For each* j, $r_{j,n} = \sqrt{n}$ *or* $r_{j,n}/\sqrt{n} \to 0$, $r_n = \min_{1\leq j\leq q} r_{j,n} \to \infty$. $\frac{1}{n^2}\sum_j\sum_k \tilde{\mathbf{f}}\left(\mathbf{X}_j\right) D_{jk}\tilde{\mathbf{f}}\left(\mathbf{X}_k\right)'$ *is finite and positive definite, where* $\tilde{\mathbf{f}}\left(\mathbf{X}_t\right) = \mathbf{f}\left(\mathbf{X}_t\right) - \frac{1}{n}\sum_{j=1}^n \mathbf{f}\left(\mathbf{X}_j\right)$.

Assumption 3 *There exists a constant* C, *such that* $E\left(\varepsilon_t|\mathbf{X}_t\right) = 0, E\left(\boldsymbol{\eta}_t|\mathbf{X}_t\right) = 0$, $E\left(\varepsilon_t^2|\mathbf{X}_t\right) < C, E\left(\left\|\boldsymbol{\eta}_t\right\|^2|\mathbf{X}_t\right) < C$, $Var\left(\left(\varepsilon_t, \boldsymbol{\eta}_t'\right)'|\mathbf{X}_t\right) = diag\left(\Omega_t^*, 0\right)$, *and* $\vartheta_{\min}\left(\sum_{t=1}^n \Omega_t^*\right) \geq 1/C$, *a.s.*

Assumption 1 allows for the i.i.d. observations. Potentially, we can extend this to allow for weakly dependent time series processes.

Assumption 2 is quite similar to Assumption 2 in Hausman et al. (2012), allowing linear combinations of $\boldsymbol{\beta}$ to have different degrees of identifications. It accommodates IV regressions involving included exogenous variables. For example, consider an IV regression with one endogenous variable, one included exogenous variable, and one instrumental variable.

$$y_t = \alpha_0 + +\beta_{01}Z_{t1} + \beta_{02}Y_{t1} + \varepsilon_t,$$

where Z_{t1} is the included exogenous variables. Hence, $\mathbf{Y}_t = (Z_{t1}, Y_{t1})'$, $\mathbf{X}_t = (Z_{t1}, X_{t1})'$. Let the reduced form be partitioned conformably with $\boldsymbol{\beta} = (\beta_{01}, \beta_{02})'$. As

$$\begin{aligned} E\left(\mathbf{Y}_t|\mathbf{X}_t\right) &= \left(Z_{t1}, \pi_1 Z_{t1} + \frac{r_{2n}}{\sqrt{n}} f_2\left(X_t\right)\right)' \\ &= \begin{bmatrix} 1 & 0 \\ \pi_1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{n}/\sqrt{n} & 0 \\ 0 & r_{2n}/\sqrt{n} \end{bmatrix} \begin{bmatrix} Z_{t1} \\ f_2\left(X_t\right) \end{bmatrix}. \end{aligned}$$

This reduced form is specified in Assumption 2 with

$$\tilde{\mathbf{R}}_n = \begin{bmatrix} 1 & 0 \\ \pi_1 & 1 \end{bmatrix}, r_{1n} = \sqrt{n}, \mathbf{f}(\mathbf{X}_t) = \begin{bmatrix} Z_{t1} \\ f_2(X_t) \end{bmatrix}.$$

Although we do not generally require $\mathbf{f}(\mathbf{X}_t)$ to be known and linear in included exogenous variables, this study has been unable to simplify Assumption 2 while including such basic situations. The positive definiteness of $\frac{1}{n^2}\sum_j \sum_k \tilde{\mathbf{f}}(\mathbf{X}_j) D_{jk} \tilde{\mathbf{f}}(\mathbf{X}_k)'$ implies its minimum eigenvalue is positive. Further, it is noted that the rates of decay to zero are slower than $\sqrt{n}$, this has been labeled as nearly-weak identification or semi-strong identification by some previous studies. This study adopts the "many weak instruments" tag, following Hansen et al. (2008), Newey and Windmeijer (2009), and Hausman et al. (2012).

Assumption 3 is similar to Assumption 3 in Hausman et al. (2012), requiring bounded second conditional moments of disturbances and uniform nonsingularity of the variance of the reduced form of disturbances.

Theorem 4.1 establishes the consistency for WCIV and WCIVF.

Theorem 4.1 *Under Assumptions 1-3, for* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ *or* $\hat{\boldsymbol{\beta}}_{WCIVF}$*,* $\hat{\alpha} = \hat{\alpha}_{WCIV}$ *or* $\hat{\alpha}_{WCIVF}$

$$\mathbf{R}_n' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0,$$
$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}, \; \hat{\alpha} \xrightarrow{p} \alpha_0.$$

Proof. See the Appendix. ■

To discuss asymptotic normality, some additional assumptions are required.

Assumption 4 *There exists a constant* $C > 0$*, such that* $E\left(\varepsilon_t^4 | \mathbf{X}_t\right) < C, E\left(\|\boldsymbol{\eta}_t\|^4 | \mathbf{X}_t\right) < C$ *a.s.* $E\|W_t\|^4 < \infty$.

We state the asymptotic normality theorem.

Theorem 4.2 *Under Assumptions 1-4, for* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ *or* $\hat{\boldsymbol{\beta}}_{WCIVF}$*,*

$$\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\boldsymbol{\Omega}(\boldsymbol{\theta}_0)\sqrt{n}\mathbf{R}_n^{-1}\right)^{-1/2}\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\boldsymbol{\Upsilon}\mathbf{R}_n^{-1}\sqrt{n}\right)\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} N(0, \mathbf{I}_p),$$

where

$$\boldsymbol{\Upsilon} = -E\left[\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right],$$

$$\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = \mathbf{S}_1(\boldsymbol{\theta}_0) + \mathbf{S}_2(\boldsymbol{\theta}_0) - \mathbf{S}_3(\boldsymbol{\theta}_0) - \mathbf{S}_3(\boldsymbol{\theta}_0)'$$

in which

$$\mathbf{S}_1(\boldsymbol{\theta}_0) = E\left(\left(y_t^{++} - \alpha_0 - \boldsymbol{\beta}_0'\mathbf{Y}_t^{++}\right)^2 (\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left\|\mathbf{X}_t - \mathbf{X}_t^{++}\right\| \left(\mathbf{Y}_t^{+} - \boldsymbol{\mu}_Y\right)' \left\|\mathbf{X}_t^{+} - \mathbf{X}_t^{++}\right\|\right),$$

$$\mathbf{S}_2(\boldsymbol{\theta}_0) = E\left[\left(y_t - \alpha_0 - \boldsymbol{\beta}_0'\mathbf{Y}_t\right)^2\right] E\left((\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right) E\left((\mathbf{Y}_t - \boldsymbol{\mu}_Y)'\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right),$$

$$\mathbf{S}_3(\boldsymbol{\theta}_0) = E\left(\left(y_t^{+} - \alpha_0 - \boldsymbol{\beta}_0'\mathbf{Y}_t^{+}\right)^2 (\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right) E\left((\mathbf{Y}_t - \boldsymbol{\mu}_Y)'\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right).$$

Proof. See the Appendix. ■ In this theorem, we establish the asymptotic normality for $\hat{\boldsymbol{\beta}}_{WCIV}$ and $\hat{\boldsymbol{\beta}}_{WCIVF}$. We regard α_0 as a nuisance parameter and do not pursue its asymptotic distribution.

In the following theorem, we establish the validity of the Wald test statistic for parameter inference regarding $\boldsymbol{\beta}_0$.

Theorem 4.3 *Under Assumptions 1-4, if* $\mathbf{g}(\cdot)$ *is continuously differentiable twice and* $\mathbf{G}(\boldsymbol{\beta}_0)$ *is of full rank, for testing the null (5), considering* $\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right) = \left(\hat{\boldsymbol{\theta}}_{WCIV}, \hat{\lambda}_{WCIV}\right)$ *or* $\left(\hat{\boldsymbol{\theta}}_{WCIVF}, \hat{\lambda}_{WCIVF}\right)$*,*

$$\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}\right) \xrightarrow{p} \boldsymbol{\Omega}(\boldsymbol{\theta}_0),$$

$$\hat{\boldsymbol{\Upsilon}}\left(\hat{\boldsymbol{\lambda}}\right) \xrightarrow{p} \boldsymbol{\Upsilon},$$

$$W_n\left(\hat{\boldsymbol{\theta}}\right) \xrightarrow{d} \chi_m^2.$$

Proof. See the Appendix. ■

This theorem demonstrates that under the null, $W_n\left(\hat{\boldsymbol{\theta}}\right)$ has a convenient chi-squared distribution asymptotically, despite the fact that the degrees of identification are unknown. An important implication of the Wald test statistic is that, without the knowledge of the degrees of weak identification, a large sample inference can be conducted in the usual way. In particular, we can obtain the t-statistic by treating $\hat{\boldsymbol{\beta}}$ as if it were normally distributed with mean $\boldsymbol{\beta}_0$ and variance $\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)/n$. Under the null, the t-statistic $\left(\hat{\beta}_j - \beta_{0j}\right)/\sqrt{\hat{V}_{jj}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)/n}$ will be asymptotically normal.Our Monte Carlo simulations show that t-statistics have excellent finite sample properties for a wide range of scenarios. In the application, we report $\sqrt{\hat{V}_{jj}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)/n}$, as if it is the conventional standard deviation.

The estimation efficiency of an estimator is a highly desired property. Under the standard asymptotic framework, the estimation efficiency of IV estimators can be achieved by utilizing an increasing number of instruments. There is some discussion on estimation efficiency in the literature on many instruments, see, for example, Hahn (2002), Anderson et al. (2010), and Kunitomo (2012). However, these theoretical results are quite limited. In many weak instruments asymptotics, an increasing number of instruments are required to ensure estimation consistency. The ratio between the number of IVs and the sample size does not necessarily align with the one required by an efficient IV estimation. Additionally, the estimation efficiency of IV estimators involves optimal weighting matrix, which is difficult to be estimated accurately under many instruments and heteroskasticity of unknown form. Regarding WCIV and WMD (WCIVF and WMDF), intuition suggests WCIV (WCIVF) is more efficient than WMD (WMDF) due to the specialty of the nonintegrable weighting function. However the theoretical validation is quite challenging and beyond the scope of this study. This study resorts to Monte Carlo simulations to evaluate the accuracy of the asymptotic approximations and compare the performance of these competitive estimators.

5 Monte Carlo Evidence

In this section, we evaluate the finite sample performance of WCIV and WCIVF, and compare it with that of WMD, WMDF, and HFUL. To describe HFUL, denote $\mathbf{X}_t^r = \left(X_{1,t}^r, ..., X_{q,t}^r\right)'$, where r is a positive integer. The instruments include a constant and pairwise instruments

$$\left(\mathbf{X}_t', \left(\mathbf{X}_t^2\right)', \left(\mathbf{X}_t^3\right)', \left(\mathbf{X}_t^4\right)', \mathbf{X}_t' d_1, ..., \mathbf{X}_t' d_{L-4}\right)',$$

where $d_l \in \{0,1\}$ and $\Pr(d_l = 1) = 1/2$. We consider $L = 1$, 4, or 9, that is, when $L = 1$, the instruments are $(1, \mathbf{X}_t')'$; when $L = 4$, $(1, \mathbf{X}_t', \left(\mathbf{X}_t^2\right)', \left(\mathbf{X}_t^3\right)', \left(\mathbf{X}_t^4\right)')'$; when $L = 9$,

$$(1, \mathbf{X}_t', \left(\mathbf{X}_t^2\right)', \left(\mathbf{X}_t^3\right)', \left(\mathbf{X}_t^4\right)', \mathbf{X}_t' d_1, ..., \mathbf{X}_t' d_5)'.$$

We denote these HFUL estimators as HFUL1, HFUL4, and HFUL9, respectively. The comparisons are in terms of median bias, range between the 0.05 and 0.95 quantiles, and empirical rejection frequencies for t-statistics at the 5% nominal level of the estimators. The number of Monte Carlo simulations is $10,000$.

5.1 Setup 1

Consider following linear models M_1-M_3, such that

$$M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t,\ Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t,$$

$$M_2 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t,\ Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t}^2 + \eta_t,$$

$$M_3 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t,\ Y_t = 1\left\{ \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t > 0 \right\},$$

where $1\{\cdot\}$ denotes the indicator function. To mimic empirical situations, allow $\mathbf{X}_t = (X_{1,t}, ..., X_{q,t})'$ to follow

$$X_{j,t} = \frac{e_{0,t} + e_{j,t}}{\sqrt{2}},\ j = 1, ..., q,$$

where $(e_{0,t}, e_{1,t}, ..., e_{q,t})' \sim i.i.d.N\left(0, \mathbf{I}_{q+1}\right)$. By construction, the correlation coefficient between $X_{j,t}$ and $X_{k,t}$ for $j \neq k$ is 0.5 owing to the presence of the common shocks $e_{0,t}$. ε_t is allowed to be heteroskedastic as

$$\varepsilon_t = \rho\eta_t + \sqrt{\frac{1-\rho^2}{\phi^2 + (0.86)^4}} \left(\phi\eta_{1,t} + 0.86\eta_{2,t}\right), \eta_{1,t} \sim N\left(0, X_{1,t}^2\right), \eta_{2,t} \sim N\left(0, 0.86^2\right),$$

where $\eta_{1,t}$ and $\eta_{2,t}$ are independent of η_t. Hausman et al. (2012) show that this design causes LIML to be inconsistent when $\phi \neq 0$. In M_1, $E\left(Y_t|\mathbf{X}_t\right)$ is linear. In M_2 and M_3, $E\left(Y_t|\mathbf{X}_t\right)$ is nonlinear. We set $\alpha_0 = \beta_0 = 0$ without loss of generality and consider a sample size of $n = 250$, $c = 10$ and $\rho = 0.6$. Further, we consider $q = 3, 10, 15$, $\phi = 0$, 0.5. When $\phi = 0$, ε_t is homoskedastic, $\phi = 0.5$, ε_t heteroskedastic.

In Tables 1–3, we report the simulation results on β_0 for WCIV, WCIVF($C = 1$), WMD, WMDF($C = 1$), HFUL1($C = 1$), HFUL4($C = 1$) and HFUL9($C = 1$). The main features of the results are as follows:

1. For M_1, when $q = 3$, HFUL1 has the best performance in terms of the range between the 0.05 and 0.95 quantiles (DecR), while HFUL4 and HFUL9 are much more dispersed. However, for $q = 10$ and 15, WCIV and WCIVF outperform HFUL1, HFUL4, and HFUL9

regarding DecR when $\phi = 0$ or 0.5. Additionally, WCIV and WCIVF are almost median unbiased for all cases, whereas HFUL1, HFUL4, and HFUL9 show relatively large median biases, consistent with the results of simulations in Hausman et al. (2012). With regard to the empirical properties of the t-statistics, both WCIV and WCIVF have accurate empirical sizes, whereas HFUL1 is undersized, and HFUL9 is oversized, especially for high-dimensional cases. WMD and WMDF exhibit comparable features to WCIV and WCIVF in terms of median biases and empirical properties of the t-statistics but have substantially larger DecR, as expected. Furthermore, while both WCIVF and WCIV perform similarly, WMDF outperforms WMD in terms of DecR but performs worse than WMD in terms of median biases and properties of the t-statistics, particularly for high-dimensional cases.

2. For M_2, HFUL1 is severely median biased and dispersed, while HFUL4 and HFUL9 is much less biased and less dispersed, as the linear instruments employed in HFUL1 cannot approximate the nonlinear reduced form sufficiently. However, both WCIV and WCIVF are almost median unbiased, with empirical rejection frequencies for the t-statistics well controlled. In terms of DecR, WCIV and WCIVF outperform WMD and WMDF substantially, and better than HFULs except for HFUL4 in the case of $q = 3$.

3. For M_3, HFUL1, HFUL4, and HFUL9 are all heavily median biased, especially when $q = 3$, while WCIV and WCIVF are almost median unbiased in all cases. So are WMD and WMDF when q is small. When q is large, however, it appears that WMDF worsens in terms of median bias, whereas it is less dispersed than WMD. In terms of DecR, WCIV and WCIVF outperform WMD and WMDF substantially in all cases, and better than HFULs except for HFUL1 in the case of $q = 3$.

In summary, we conclude that WCIV and WCIVF have exceptional finite-sample properties in the context of Setup 1. They exhibit almost median unbiasedness in all cases, and their empirical rejection frequencies of the t-statistics are close to the nominal value. They are considerably less dispersed than WMD and WMDF in all cases. In comparison with HFUL, both WCIV and WCIVF exhibit less dispersion in numerous cases, particularly for nonlinear reduced forms and large values of q. Furthermore, HFUL is generally more biased than WCIV and WCIVF. Additionally, the finite-sample properties of HFUL are significantly sensitive to the number of selected instruments, particularly when the reduced forms are nonlinear. These findings demonstrate that

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$\phi = 0$					$q = 3$			
	Med	0.0015	0.0016	0.0017	0.0047	0.0330	0.0339	0.0464
	DecR	0.8347	0.8341	1.0607	1.0452	0.7504	0.8806	1.1013
	Rej	0.0537	0.0537	0.0547	0.0554	0.0336	0.0510	0.0797
					$q = 10$			
	Med	0.0016	0.0018	0.0015	0.0096	0.0111	0.0151	0.0209
	DecR	0.4722	0.4717	0.7321	0.7052	0.4768	0.6338	0.9444
	Rej	0.0513	0.0514	0.0458	0.0506	0.0114	0.0308	0.0740
					$q = 15$			
	Med	0.0000	0.0003	-0.0004	0.0337	0.0084	0.0126	0.0293
	DecR	0.3782	0.3780	0.7192	0.6159	0.3877	0.5617	0.9736
	Rej	0.0472	0.0474	0.0507	0.0664	0.0040	0.0197	0.0741
$\phi = 0.5$					$q = 3$			
	Med	-0.0031	-0.0030	-0.0061	-0.0034	0.0361	0.0400	0.0524
	DecR	0.9336	0.9337	1.1291	1.1103	0.8453	0.9939	1.2257
	Rej	0.0503	0.0503	0.05	0.0509	0.0362	0.0574	0.0882
					$q = 10$			
	Med	-0.0017	-0.0015	-0.0012	0.0070	0.0135	0.0172	0.0224
	DecR	0.4993	0.4989	0.7219	0.6993	0.5141	0.6510	0.9754
	Rej	0.0458	0.0458	0.0507	0.0538	0.0144	0.0337	0.0738
					$q = 15$			
	Med	0.0002	0.0005	-0.0018	0.0317	0.0111	0.0141	0.0228
	DecR	0.4054	0.4048	0.6831	0.5891	0.4186	0.5765	0.9542
	Rej	0.0500	0.0500	0.0487	0.0632	0.0080	0.0285	0.0768

Table 1: Linear IV model $M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t$, $Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t$. Median bias (Med), the range between the 0.05 and 0.95 quantiles (DecR), and empirical rejection frequencies for t-statistics at 5% nominal level (Rej) are reported.

HFUL may provide misleading estimates when the reduced forms are not well-approximated using linear combinations of the selected instruments.

5.2 Setup 2

We consider alternative linear models M_4-M_6, which are similar to Antoine and Lavergne (2014), such that

$$M_4 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_t,\ Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \eta_t,$$

$$M_5 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_t,\ Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \exp(0.5 + 0.5X_{1,t,})\eta_t,$$

$$M_6 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_t,\ Y_t = \exp\left(\frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t}\right) + \eta_t.$$

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$\phi = 0$					$q = 3$			
	Med	0.0000	0.0009	-0.0020	0.0028	0.5617	0.0267	0.0272
	DecR	1.0401	1.0326	1.2310	1.1989	1.5104	0.7683	0.9729
	Rej	0.0398	0.0398	0.0512	0.0516	0.5472	0.0436	0.0701
					$q = 10$			
	Med	0.0000	0.0029	0.0016	0.0163	0.4764	0.0137	0.0281
	DecR	0.6050	0.5941	1.0552	0.9506	1.9397	0.6783	1.0975
	Rej	0.0373	0.0382	0.0504	0.0558	0.5252	0.0256	0.0719
					$q = 15$			
	Med	-0.0009	0.0012	-0.0026	0.0576	0.4480	0.0142	0.0470
	DecR	0.4831	0.4772	1.1197	0.7513	2.0637	0.6847	1.2875
	Rej	0.0366	0.0379	0.0574	0.0808	0.5118	0.0260	0.0804
$\phi = 0.5$					$q = 3$			
	Med	-0.0189	-0.0165	-0.0118	-0.0069	0.5668	0.0317	0.0379
	DecR	1.1754	1.1458	1.2764	1.2380	1.5758	0.8506	1.0585
	Rej	0.0359	0.0362	0.0474	0.0482	0.5313	0.0555	0.0808
					$q = 10$			
	Med	-0.0092	-0.0059	-0.0107	0.0048	0.5072	0.0178	0.0310
	DecR	0.6711	0.6615	1.0626	0.9508	1.9612	0.7379	1.1046
	Rej	0.0351	0.0366	0.0466	0.0519	0.5471	0.0345	0.0757
					$q = 15$			
	Med	-0.0030	-0.0003	-0.0030	0.0590	0.4479	0.0168	0.0399
	DecR	0.5380	0.5282	1.0702	0.7366	2.0453	0.7144	1.2552
	Rej	0.0408	0.0415	0.055	0.0780	0.5172	0.0302	0.085

Table 2: Linear IV model $M_2 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t$, $Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t}^2 + \eta_t$. Median bias (Med), the range between the 0.05 and 0.95 quantiles (DecR), and empirical rejection frequencies for t-statistics at 5% nominal level (Rej) are reported.

Note that heteroskedasticity in model disturbances is allowed for. In all models ε_t and η_t follow a joint normal distribution with a covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. $\mathbf{X}_t = (X_{1,t}, ..., X_{q,t})'$ follows the same process as in Setup 1. The reduced form in M_4 is a linear model with homoskedastic errors; the reduced form in M_5 is a linear model with heteroskedastic errors; while the reduced form in M_6 is nonlinear. We set $\alpha_0 = \beta_0 = 0$ again. In the simulations, we values $c = 4$, 8, $\rho = 0.8$ and $n = 250$. Clearly, when $c = 4$, the degree of weak identification is more severe. We consider $q = 4, 8$ and 16 to check the finite sample properties of estimators under different dimensions of conditioning variables.

Tables 4–6 report the simulation results of β_0 for WCIV, WCIVF($C = 1$), WMD, WMDF($C = 1$), HFUL1($C = 1$), HFUL4($C = 1$), and HFUL9($C = 1$). The general conclusions are similar to those presented in Setup 1. That is, WCIV and WCIVF have excellent finite sample properties,

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$\phi = 0$					$q = 3$			
	Med	-0.0062	-0.0057	0.0006	0.0091	0.0843	0.1167	0.1554
	DecR	2.3123	2.3006	3.1884	3.0489	1.9550	2.5684	3.5802
	Rej	0.0442	0.0442	0.0457	0.0465	0.0233	0.0478	0.0829
					$q = 10$			
	Med	-0.0095	-0.0090	-0.0046	0.0175	0.0369	0.0455	0.1031
	DecR	1.2964	1.2948	2.3624	2.2003	1.3876	2.0770	3.4899
	Rej	0.0419	0.0420	0.0462	0.0483	0.0094	0.0327	0.0726
					$q = 15$			
	Med	-0.0026	-0.0017	-0.0114	0.0759	0.0248	0.0412	0.1182
	DecR	1.1017	1.1006	2.3307	1.8374	1.2120	1.9415	3.7257
	Rej	0.0483	0.0487	0.0478	0.0606	0.0046	0.0309	0.0800
$\phi = 0.5$					$q = 3$			
	Med	-0.0028	-0.0023	-0.0121	-0.0043	0.1153	0.1288	0.1585
	DecR	2.5261	2.5255	3.3285	3.2036	2.1789	2.8813	3.8007
	Rej	0.0417	0.0418	0.0441	0.0445	0.0302	0.0562	0.0855
					$q = 10$			
	Med	-0.0011	-0.0005	0.0020	0.0231	0.0474	0.0617	0.0995
	DecR	1.3954	1.3937	2.2599	2.1132	1.4886	2.1757	3.5166
	Rej	0.0420	0.0420	0.0437	0.0465	0.0142	0.0402	0.0783
					$q = 15$			
	Med	-0.0055	-0.0049	-0.0116	0.0761	0.0298	0.0315	0.1004
	DecR	1.1689	1.168	2.3200	1.7693	1.3057	2.0461	3.8185
	Rej	0.0459	0.046	0.0471	0.0585	0.0075	0.0328	0.0806

Table 3: Linear IV model $M_3 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t$, $Y_t = 1\left\{\sqrt{\frac{c/q}{n}}\sum_{j=1}^{q} X_{j,t} + \eta_t > 0\right\}$. Median bias (Med), the range between the 0.05 and 0.95 quantiles (DecR), and the empirical rejection frequencies for t-statistics at the 5% nominal level (Rej) are reported.

outperforming other alternatives, especially when the q values are large. On the other hand, when the weak identification is severe, HFUL has very poor finite sample properties. Notably HFUL is heavily biased in the case of M_5.

6 Application to Estimating the EIS in Consumption

In this section, WCIV and WCIVF are applied to estimate the EIS in consumption for macro datasets from the US. The EIS in consumption is a parameter of central importance in macroeconomics and finance as it measures how much consumers change their expected consumption growth rate in response to changes in the expected return on any asset. For example, King and Rebelo (1990) demonstrates that EIS is the key parameter in a simple neoclassical model of endogenous growth, which involves taxation. In the consumption based asset pricing models, the

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$c = 4$					$q = 4$			
	Med	-0.0101	-0.0098	-0.0238	-0.0186	0.0551	0.0629	0.0719
	DecR	1.1091	1.1065	1.3296	1.2916	0.9853	1.1049	1.2708
	Rej	0.0582	0.0584	0.0595	0.0608	0.066	0.0887	0.1168
					$q = 8$			
	Med	-0.0044	-0.0041	-0.0123	-0.0029	0.0312	0.0381	0.0501
	DecR	0.7490	0.7479	0.9619	0.923	0.7440	0.8893	1.1267
	Rej	0.0506	0.0506	0.0548	0.057	0.0446	0.0671	0.0958
					$q = 16$			
	Med	-0.0036	-0.0030	-0.0099	0.0718	0.0173	0.0236	0.0482
	DecR	0.5232	0.5222	0.8615	0.5967	0.5459	0.6845	1.0630
	Rej	0.0473	0.0475	0.056	0.0913	0.0242	0.0484	0.0983
$c = 8$					$q = 4$			
	Med	-0.0048	-0.0047	-0.0121	-0.0099	0.0274	0.0306	0.0301
	DecR	0.7233	0.7229	0.7913	0.7831	0.7120	0.7642	0.8497
	Rej	0.0511	0.0513	0.0516	0.0525	0.0608	0.0693	0.0836
					$q = 8$			
	Med	-0.0023	-0.0022	-0.0068	-0.0020	0.0164	0.0203	0.0237
	DecR	0.5104	0.5097	0.6007	0.5920	0.5195	0.5636	0.6661
	Rej	0.0495	0.0496	0.0465	0.0479	0.0439	0.0538	0.0721
					$q = 16$			
	Med	-0.0019	-0.0017	-0.0058	0.0364	0.0080	0.0106	0.0202
	DecR	0.3611	0.3610	0.5276	0.4518	0.3752	0.4241	0.5966
	Rej	0.0511	0.0511	0.0460	0.0710	0.0244	0.0361	0.0732

Table 4: Linear IV model $M_4 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_t,\ Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \eta_t$. Median bias (Med), the range between the 0.05 and 0.95 quantiles (DecR), and empirical rejection frequencies for t-statistics at 5% nominal level (Rej) are reported.

EIS determines the optimal consumption rule, as observed in Campbell and Viceira (1999).

Therefore, EIS is a key input parameter in many macroeconomic or financial model calibrations. In recent years, the EIS has been set to be quite large in many cases, reflecting the general view among macroeconomists today that a high EIS is more consistent with the stylized facts of macroeconomic dynamics. For example, Bansal and Yaron (2004) choose an EIS value as large as 1.5, while Barro (2009), Ai (2010), and Colacito and Croce (2011) set the EIS value to 2. However, to date, empirical estimation results based on macro data sets have provided limited support to this view.[2] Early literature, such as Hansen and Singleton (1983), has suggested EIS values as high as one. However Hall (1988) argues that they do not consider time aggregation

[2]At the micro data level, there is some evidence of a high EIS value. For example, Attanasio and Weber (1993) find higher values for using disaggregated cohort-level consumption data; Vissing-Jorgensen (2002), using household data, records a higher EIS value among asset market participants. However, these results do not directly support the large EIS values observed in macro model calibrations because they are based on aggregate macro data.

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$c = 4$					$q = 4$			
	Med	0.0334	0.0343	0.0432	0.0518	0.1375	0.1841	0.2064
	DecR	1.8895	1.7986	2.8125	1.7637	0.8177	1.1188	1.1978
	Rej	0.0835	0.0836	0.1086	0.1119	0.1311	0.2218	0.2415
					$q = 8$			
	Med	0.0054	0.0065	0.0262	0.0423	0.0781	0.1337	0.1765
	DecR	1.0639	1.0260	2.1302	1.2105	0.7390	1.1126	1.2096
	Rej	0.071	0.0711	0.1046	0.1107	0.0955	0.1938	0.2178
					$q = 16$			
	Med	-0.0008	0.0000	0.0393	0.1381	0.0366	0.0940	0.1624
	DecR	0.6651	0.6531	2.3379	0.5004	0.6126	1.0625	1.1677
	Rej	0.0656	0.0664	0.1198	0.1950	0.0723	0.1905	0.2161
$c = 8$					$q = 4$			
	Med	0.0050	0.0053	0.0040	0.0095	0.0687	0.1023	0.1257
	DecR	0.9909	0.9761	1.2214	1.0523	0.6650	1.0186	1.1128
	Rej	0.0699	0.0700	0.0831	0.0850	0.0908	0.1689	0.1872
					$q = 8$			
	Med	-0.0006	0.0000	-0.0005	0.0093	0.0324	0.0547	0.0887
	DecR	0.6195	0.6128	0.8948	0.7338	0.5392	0.7956	1.0709
	Rej	0.0619	0.062	0.0784	0.0819	0.0632	0.127	0.1671
					$q = 16$			
	Med	-0.0010	-0.0006	0.0013	0.0749	0.0140	0.0321	0.0799
	DecR	0.4081	0.4058	0.8885	0.3698	0.4309	0.6528	0.9985
	Rej	0.058	0.0587	0.0842	0.1386	0.0381	0.114	0.1557

Table 5: Linear IV model $M_5 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_t$, $Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \exp(0.5 + 0.5X_{1,t,})\eta_t$. Median bias (Med), the range between the 0.05 and 0.95 quantiles (DecR), and the empirical rejection frequencies for t-statistics at the 5% nominal level (Rej) are reported.

problem of the data appropriately, and the employed instruments are problematic. When valid instruments are employed, Hall (1988) finds that the TSLS estimates of the EIS for the US are unlikely to be much higher than 0.1 and may well be 0. Yogo (2004) points out that these misleading results may be attributed to weak instruments. Yogo (2004) and Ascari et al. (2021) employ weak-instrument-robust inference procedures on macro data sets, following Staiger and Stock (1997), Kleibergen (2002), Moreira (2003) and Kleibergen (2005); however, they reach similar conclusions as in Hall (1988). It should be noted that these estimation and inference procedures only employ a fixed number of instruments. Therefore, it would be of great interest to re-examine the EIS estimation with the estimation procedures in the literature on many weak instruments.

To derive the estimable log-linearized Euler equation, we consider a basic consumption-based asset pricing model with the Epstein-Zin utility function. Let δ be the subjective discount factor,

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$c = 4$					$q = 4$			
	Med	-0.0101	-0.0099	-0.0220	-0.0177	0.0533	0.0550	0.0641
	DecR	1.0618	1.0613	1.2879	1.2584	0.9502	1.0748	1.2412
	Rej	0.0556	0.056	0.0597	0.0605	0.0668	0.0839	0.1093
					$q = 8$			
	Med	0.0005	0.0009	-0.0039	0.0033	0.0331	0.0339	0.0380
	DecR	0.7052	0.7036	0.9136	0.8756	0.7023	0.8054	1.0154
	Rej	0.0541	0.0547	0.0586	0.0619	0.0455	0.0679	0.0915
					$q = 16$			
	Med	0.0011	0.0016	-0.0035	0.0642	0.0168	0.0206	0.0352
	DecR	0.4648	0.4646	0.7355	0.5478	0.4931	0.5847	0.8606
	Rej	0.0488	0.049	0.0555	0.0875	0.0231	0.0437	0.0872
$c = 8$					$q = 4$			
	Med	-0.0051	-0.0050	-0.0118	-0.0096	0.0247	0.0228	0.0239
	DecR	0.6652	0.6651	0.7349	0.7287	0.6644	0.6928	0.7671
	Rej	0.0474	0.0475	0.0495	0.0502	0.061	0.064	0.0787
					$q = 8$			
	Med	0.0007	0.0008	-0.0015	0.0024	0.0155	0.0168	0.0164
	DecR	0.4504	0.4502	0.5311	0.5229	0.4627	0.4816	0.5492
	Rej	0.0531	0.0534	0.0495	0.0513	0.0495	0.0565	0.0696
					$q = 16$			
	Med	0.0005	0.0007	-0.0014	0.0283	0.0070	0.0098	0.0135
	DecR	0.2884	0.2883	0.4147	0.3678	0.3078	0.3228	0.4167
	Rej	0.0503	0.0507	0.0485	0.068	0.0293	0.0349	0.0693

Table 6: Linear IV model $M_6 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_t$, $Y_t = \exp\left(\frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t}\right) + \eta_t$. Median bias (Med), the range between the 0.05 and 0.95 quantiles (DecR), and the empirical rejection frequencies for t-statistics at the 5% nominal level (Rej) are reported.

γ be the coefficient of relative risk aversion, and $\theta = (1-\gamma)/(1-1/\psi)$. Following Epstein and Zin (1989) and Epstein and Zin (1991), the objective utility function is defined recursively by

$$U_t = \left[(1-\delta) C_t^{(1-\gamma)/\theta} + \delta \left(E_t U_{t+1}^{1-\gamma}\right)^{1/\theta}\right]^{\theta/(1-\gamma)}, \tag{14}$$

where C_t is consumption at time t; E_t denotes conditional expectation $E(\cdot|\mathcal{F}_t)$, where $\mathcal{F}_t$ is the information set at time t. In the special case where $\gamma = 1/\psi$, (14) reduces to the familiar time-separable power utility model with period utility function $U(C_t) = C_t^{1-\gamma}/(1-\gamma)$. The representative household maximizes the objective function (14) subject to the intertemporal budget constraint

$$W_{t+1} = (1 + R_{w,t+1})(W_t - C_t) \tag{15}$$

where W_{t+1} is the household's wealth and $1+R_{w,t+1}$ is the gross real return on the portfolio of all invested wealth at $t+1$. Epstein and Zin (1991) show that equations (14) and (15) imply the Euler equation of the form

$$E_t\left[\left(\delta\left(\frac{C_{t+1}}{C_t}\right)^{-1/\psi}\right)^{\theta}\left(\frac{1}{1+R_{w,t+1}}\right)^{1-\theta}(1+R_{j,t+1})\right]=1 \tag{16}$$

where $1+R_{j,t+1}$ is the gross real return on asset j.

Let lowercase letters denote the logarithms of the corresponding uppercase variables (e.g., $r_{j,t+1}=\log(1+R_{j,t+1})$). By assuming that asset returns and consumption are homoskedastic and jointly log normal conditional on F_t, the Euler equation (16) can be linearized as

$$E_t\left(r_{j,t+1}-\eta_j-\frac{1}{\psi}\triangle c_{t+1}\right)=0, \tag{17}$$

where ψ is the EIS in consumption and

$$\eta_j=\eta_f-\frac{1}{2}Var\left(r_{j,t+1}-E_t r_{j,t+1}\right)+\frac{\theta}{\psi}Cov\left(r_{j,t+1}-E_t r_{j,t+1},\triangle c_{t+1}-E_t\triangle c_{t+1}\right)$$
$$+(1-\theta)\,Cov\left(r_{j,t+1}-E_t r_{j,t+1},r_{w,t+1}-E_t r_{w,t+1}\right),$$

where

$$\eta_f=-\log\delta+\frac{\theta-1}{2}Var\left(r_{w,t+1}-E_t r_{w,t+1}\right)-\frac{\theta}{2\psi^2}Var\left(\triangle c_{t+1}-E_t\triangle c_{t+1}\right).$$

If asset returns and consumption are conditionally heteroskedastic, we can still obtain a similar linearized Euler equation; however, $r_{j,t+1}-\eta_j-\frac{1}{\psi}\triangle c_{t+1}$ is now heteroskedastic; see Yogo (2004) for a more detailed discussion.

Based on (17), Hall (1988), Campbell (2003), Yogo (2004), and Ascari et al. (2021), among others, have used an instrumental variable regression approach to estimate EIS. Normally one would choose a vector $\mathbf{X}_t$, which is a subset of the information set. By the law of iterated expectations, we get

$$E\left[r_{j,t+1}-\eta_j-1/\psi\triangle c_{t+1}|\mathbf{X}_t\right]=0, \tag{18}$$

or its reversed form

$$E\left[\triangle c_{t+1}-\alpha_j-\psi r_{j,t+1}|\mathbf{X}_t\right]=0. \tag{19}$$

6.1 The US Quarterly Data in Ascari et al. (2021)

We utilize the dataset from Ascari et al. (2021) which includes quarterly data on equity markets at an aggregate level, and macroeconomic variables from Q4 1955 to Q1 2018. For the nominal interest rate i_t, this analysis employs three-month treasury bill rate; for the nominal stock market return s_t, the S&P 500 return is employed. For c_t, the log of real consumption of nondurables and services is used, following Campbell and Mankiw (1989) and Yogo (2004). The inflation rate π_t is determined based on the deflator that corresponds to the consumption of nondurables and services. Additional details regarding the data sources and transformation techniques can be found in the supplementary appendix of Ascari et al. (2021).

As per Ascari et al. (2021), the ex-post real interest rate $i_t - \pi_{t+1}$ and the ex-post real stock return $st - \pi_{t+1}$ are considered for the analysis. In the empirical analysis, the EIS is estimated using the real interest rate as the asset return. [3] $\mathbf{X}_t$ comprises lag terms of the real interest rate, real stock return, consumption growth and the first-difference of the log dividend-price ratio ($\triangle dp_t$).[4] It is worthwhile mentioning that the first-difference of the log dividend-price ratio is considered instead of the log dividend-price ratio, due to its non-stationary nature. [5] Specifically, to estimate (18) and (19), we use $i_t - \pi_{t+1}$, $s_t - \pi_{t+1}$, Δc_t, and Δdp_t from the first lag up to the third lag. Thus they are at least lagged twice to avoid the data aggregation issue described in Hall (1988). As a comparison, the estimates obtained using alternative estimation procedures are also reported. For HFUL, the instruments include a constant and pairwise instruments $\left(\mathbf{X}_t', \left(\mathbf{X}_t^2\right)', \mathbf{X}_t' d_1, ..., \mathbf{X}_t' d_{L-2}\right)'$, where $d_l \in \{0, 1\}$, $\Pr(d_l = 1) = 1/2$. We consider $L = 1$, 2 or 6, that is, when $L = 1$, the instruments are $(1, \mathbf{X}_t')'$; when $L = 2$, $(1, \mathbf{X}_t', \left(\mathbf{X}_t^2\right)')'$; when $L = 6$, $(1, \mathbf{X}_t', \left(\mathbf{X}_t^2\right)', \mathbf{X}_t' d_1, ..., \mathbf{X}_t' d_4)'$. We denote these HFUL estimates as HFUL1, HFUL2, and HFUL6, respectively. We have to utilize a smaller number of instruments to avoid singular matrix problem in HFUL. We set $C = 1$ for WCIVF, WMDF and HFULs.

Table 7 presents the estimation results for $1/\psi$ and ψ. Notably, WCIV and WCIVF($C = 1$) estimates of the EIS (ψ) appear to be large, and statistically significant at the 10% significant

[3] We did not consider the stock return as the asset return, since it is harder to predict, the problem of weak instruments is more severe , as demonstrated in previous empirical studies.

[4] In Yogo (2004), lag terms of the nominal interest rate, inflation rate, consumption growth, and log dividend-price ratio are utilized, while Campbell (2003) employs lag terms of the real interest rate, real consumption growth, and log dividend-price ratio. Additionally, Beeler and Campbell (2012) use lag terms of the real interest rate, real stock return, real consumption growth, and log dividend-price ratio. In contrast, Ascari et al. (2021) examines lag terms of the real consumption growth and real interest rate.

[5] The null hypothesis that the log price-dividend ratio is a unit root is not rejected by the Phillips-Perron test at the 5% significant level. Therefore, the first-difference of log price-dividend ratio, instead of the log price-dividend ratio, is employed.

	WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL2	HFUL6	TSLS
				lags 1				
ψ	1.03	1.01	1.19	0.77	0.15	0.22	0.19	0.14
	(0.62)	(0.60)	(0.97)	(0.45)	(0.14)	(0.11)	(0.17)	(0.10)
$1/\psi$	0.98	0.97	0.84	0.67	4.77	3.95	4.03	0.62
	(0.61)	(0.61)	(0.71)	(0.51)	(4.32)	(2.05)	(2.41)	(0.31)
				lags 1 to 2				
ψ	1.15	1.14	1.27	0.90	0.21	0.23	0.19	0.17
	(0.64)	(0.63)	(0.90)	(0.49)	(0.16)	(0.16)	(0.15)	(0.09)
$1/\psi$	0.87	0.86	0.78	0.68	3.94	3.65	3.28	0.50
	(0.50)	(0.50)	(0.58)	(0.47)	(3.00)	(2.45)	(1.67)	(0.23)
				lags 1 to 3				
ψ	1.62	1.60	1.82	1.36	0.23	0.24	0.26	0.18
	(0.94)	(0.92)	(1.16)	(0.67)	(0.18)	(0.21)	(0.27)	(0.09)
$1/\psi$	0.62	0.61	0.55	0.52	3.67	3.45	2.98	0.46
	(0.34)	(0.34)	(0.33)	(0.31)	(2.81)	(2.95)	(2.55)	(0.22)

Table 7: Estimates of the EIS using real interest rate as the asset return. The quarterly data range is from Q4 1955 to Q1 2018. EIS is estimated from $E\left[\triangle c_{t+1} - \alpha - \psi r_{t+1} | \mathbf{X}_t\right] = 0$. The reciprocal of the EIS is estimated from $E\left[r_{t+1} - \mu - 1/\psi \triangle c_{t+1} | \mathbf{X}_t\right] = 0$. $\mathbf{X}_t$ comprises lag terms of the real interest rate, real stock return, consumption growth and the first-difference of the log dividend-price ratio from the first lag up to the third lag. The values in the brackets are the standard deviations of the corresponding estimates.

level. These findings hold true over model transformation, with the WCIV and WN IVF($C = 1$) estimates of $1/\psi$ align with those of ψ. Although, the WMD and WMDF($C = 1$) estimates of the EIS are comparable to those of WCIV and WCIVF($C = 1$) in some cases, the WMD estimates are not statistically significant at the 10% significant level, and the WMDF estimates are generally smaller than the WCIV and WCIVF estimates substantially. Furthermore, the WMD and WMDF estimates frequently differ substantially. For example, for the first IV set, the WMD estimate of EIS is 1.19 and statistically insignificant at the 10% significant level, while the WMDF($C = 1$) estimate is 0.77, and statistically significant at the 10% significant level. The HFUL estimates of the EIS are generally greater than the TSLS estimates, but much less than the WCIV and WCIVF($C = 1$) estimates.

6.2 The US Quarterly Data in Beeler and Campbell (2012)

To further check the robustness of the WCIV and WCIVF estimates of the EIS, an alternative quarterly data set from Beeler and Campbell (2012) is considered. The data range is from Q2 1947 to Q4 2008. The stock market data are based on monthly CRSP NYSE/AMEX Value-weighted Indices. The real interest rates and real stock returns are ex-ante. See the appendix

	WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL2	HFUL6	TSLS
				IV Set 1				
$1/\psi$	0.49	0.49	0.54	0.43	2.82	3.51	3.39	1.62
	(0.30)	(0.30)	(0.30)	(0.22)	(1.08)	(1.60)	(1.53)	(0.40)
ψ	2.03	1.98	1.86	1.16	0.32	0.24	0.25	0.33
	(1.07)	(1.02)	(0.99)	(0.41)	(0.12)	(0.11)	(0.11)	(0.13)
				IV Set 2				
$1/\psi$	0.59	0.57	0.39	0.20	2.82	3.62	3.33	1.10
	(0.50)	(0.48)	(0.39)	(0.10)	(1.16)	(1.85)	(1.69)	(0.27)
ψ	1.69	1.60	2.58	0.40	0.32	0.23	0.25	0.32
	(0.78)	(0.71)	(1.66)	(0.11)	(0.13)	(0.11)	(0.13)	(0.13)
				IV Set 3				
$1/\psi$	0.42	0.42	0.50	0.41	2.67	3.39	3.15	1.03
	(0.24)	(0.23)	(0.26)	(0.20)	(1.03)	(1.58)	(1.46)	(0.23)
ψ	2.36	2.30	2.00	1.23	0.34	0.25	0.27	0.34
	(1.33)	(1.26)	(1.07)	(0.43)	(0.13)	(0.12)	(0.13)	(0.13)

Table 8: Estimates of the EIS using real interest rate as the asset return. The data range is from Q2 1947 to Q4 2008. The EIS is estimated from $E\left[\triangle c_{t+1} - \alpha - \psi r_{t+1} | \mathbf{X}_t\right] = 0$. The reciprocal of EIS is estimated from $E\left[r_{t+1} - \mu - 1/\psi \triangle c_{t+1} | \mathbf{X}_t\right] = 0$. The first set consists of the first lag terms of real interest rate, real stock return, and consumption growth. The second set consists of the first lag terms of real interest rate, consumption growth, and first-difference of log price-dividend ratio. The third set consists of the first lag terms of real interest rate, real stock return, consumption growth, and first-difference of log price-dividend ratio. The values in the brackets are the standard deviations of the corresponding estimates.

of Beeler and Campbell (2012) for detailed description of the data, sources, and transformation used.

The EIS is estimated using real interest rate as the asset return. Three sets of $\mathbf{X}_t$ are considered. The first set consists of lag terms of real interest rate, real stock return, and consumption growth. The second set consists of the lag terms of real interest rate, consumption growth, and first-difference of log price-dividend ratio. The third set consists of real interest rate, real stock return, consumption growth, and first-difference of log price-dividend ratio. We consider the first lag terms of $\mathbf{X}_t$ in our analysis.

The empirical results are reported in Table 8. It is observed that the WCIV and WCIVF($C = 1$) estimates of the EIS are also quite large, although the data range, data structure, and $\mathbf{X}_t$ are different. Notably, for the first set, the WCIV and WCIVF($C = 1$) estimates of the EIS are approximately 2 and statistically significant at the 5% level. For the third set, the estimates are even larger than 2. In contrast, the HFUL estimates of the EIS are quite small, even less than the TSLS estimates in several cases.

In summary, we obtain large WCIV and WCIVF($C = 1$) estimates of the EIS in consumption, which well exceed one, and are statistically significant from zero. Further, these findings are robust to different sets of $\mathbf{X}_t$, model transformation, and different data structures and data ranges. These results are strikingly different from those of HFUL and lend strong support to the practices of some model calibrations, where the EIS value is set to be significantly large.

7 Conclusion

This study proposes two novel IV estimators, namely WCIV and WCIVF, utilizing a continuum of instruments and a unique nonintegrable weighting function in the minimum distance objective function of IV estimation. This study demonstrates that these estimators are consistent and asymptotically normally distributed in the face of weak instruments and heteroskedasticity of unknown form. Extensive Monte Carlo simulations reveal that they exhibit highly favorable finite sample properties under various model setups. We apply WCIV and WCIVF to estimate the EIS of consumption for macro datasets of the US. Our results show the WCIV and WCIVF estimates well exceed one and are statistically significant, which is strikingly different from the results obtained using alternative approaches.

8 Appendix

Throughout, let C denote a generic positive constant that may be different in different uses. $\sum_{j,k} = \sum_{j=1}^{n}\sum_{k=1}^{n}$. Let w.p.a.1 denote with probability approaching one.

Lemma 8.1 *For all* $\mathbf{X} \in \mathbb{R}^q$

$$\int_{R^q} \frac{1-\cos\langle \boldsymbol{\tau}, \mathbf{X}\rangle}{\|\boldsymbol{\tau}\|^{q+1}} d\boldsymbol{\tau} = c_q \|\mathbf{X}\|,$$

where

$$c_q = \frac{\pi^{(q+1)/2}}{\Gamma\left((q+1)/2\right)},$$

in which $\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt, r \neq 0, -1, -2....$ *The integrals at* 0 *and* ∞ *are meant in the principal value sense:* $lim_{\varepsilon \to 0} \int_{\mathbb{R}^q \backslash \{\varepsilon B + \varepsilon^{-1} B^c\}}$, *where* B *is the unit ball centered at* 0 *and* B^c *is the complement of* B*, and*

$$\int_{\mathbb{R}^q} \frac{\sin\left(\langle \boldsymbol{\tau}, \mathbf{X}\rangle\right)}{\|\boldsymbol{\tau}\|^{q+1}} d\boldsymbol{\tau} = 0.$$

Proof. For any $x \in \mathbb{R}$,

$$\begin{aligned} \frac{d \int_0^\infty \frac{1-\cos(xs)}{s^2} ds}{dx} &= \int_0^\infty \frac{\sin(xs)}{s} ds \\ &= \frac{\pi}{2} sgn(x), \end{aligned}$$

where $sgn(x)$ denotes the sign of x. So

$$\int_0^\infty \frac{1-\cos(xs)}{s^2} ds = \frac{\pi}{2} |x|. \tag{20}$$

By 3.3.2.3, P586, Prudnikov et al. (1986) and applying (20), we have

$$\begin{aligned} \int_{\mathbb{R}^q} \frac{1-\cos(\langle \boldsymbol{\tau}, \mathbf{X} \rangle)}{\|\boldsymbol{\tau}\|^{q+1}} d\boldsymbol{\tau} &= \frac{2\pi^{(q-1)/2}}{\Gamma\left(\frac{q-1}{2}\right)} \int_0^\pi \int_0^\infty \frac{1-\cos(\|\mathbf{X}\| s \cos u)}{s^2} ds \sin^{q-2}(u) du \\ &= \|\mathbf{X}\| \frac{2\pi^{(q+1)/2}}{\Gamma\left(\frac{q-1}{2}\right)} \int_0^\pi |\cos u| \sin^{q-2}(u) du \\ &= \|\mathbf{X}\| \frac{\pi^{(q+1)/2}}{\Gamma\left(\frac{q+1}{2}\right)}. \end{aligned}$$

$$\begin{aligned} \int_{\mathbb{R}^q} \frac{\sin(\langle \boldsymbol{\tau}, \mathbf{X} \rangle)}{\|\boldsymbol{\tau}\|^{q+1}} d\boldsymbol{\tau} &= \frac{2\pi^{(q-1)/2}}{\Gamma\left(\frac{q-1}{2}\right)} \int_0^\pi \int_0^\infty \frac{\sin(\|X\| s \cos u)}{s^2} ds \sin^{q-2}(u) du \\ &= \|\mathbf{X}\| \frac{2\pi^{(q-1)/2}}{\Gamma\left(\frac{q-1}{2}\right)} \int_0^\infty \frac{\sin(s)}{s^2} ds \int_0^\pi \cos u \sin^{q-2}(u) du \\ &= 0. \end{aligned}$$

So the proof is complete. ■

Proof of Lemma 3.1 . Under Assumption 2, $E\|\mathbf{X}_t\|^2 < \infty$,

$$\begin{aligned} E\|\mathbf{Y}_t\|^2 &= E\left\|\frac{\mathbf{R}_n \mathbf{f}(\mathbf{X}_t)}{\sqrt{n}} + \boldsymbol{\eta}_t\right\|^2 \\ &\le 2E\left\|\frac{\mathbf{R}_n \mathbf{f}(\mathbf{X}_t)}{\sqrt{n}}\right\|^2 + 2E\|\boldsymbol{\eta}_t\|^2 \\ &\le 2\left\|\frac{\mathbf{R}_n}{\sqrt{n}}\right\|^2 E\|\mathbf{f}(\mathbf{X}_t)\|^2 + 2E\|\boldsymbol{\eta}_t\|^2 \\ &< \infty, \end{aligned}$$

and, for any $\boldsymbol{\beta} \in \mathbb{R}^p$

$$E\left|y_t\right|^2 = E\left|\alpha_0 + \boldsymbol{\beta}_0' \mathbf{Y}_t + \varepsilon_t\right|^2 < \infty.$$

Note $\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right) = \cos\left(\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right) + i\sin\left(\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)$, then by Cauchy-Schwarz inequality,

$$\begin{aligned}
\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) &= \int_{\mathbb{R}^q} \left|E\left[\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)\left(\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right) - E\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right)\right]\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \\
&\leq \int_{\mathbb{R}^q} E\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)^2 E\left|\exp i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle - E\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \\
&\leq E\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)^2 E\left[\int_{\mathbb{R}^q} 1 - \cos\left(\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right) \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right)\right] \\
&\leq E\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)^2 E\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\| \\
&< \infty.
\end{aligned}$$

Now

$$\begin{aligned}
\left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2 &= E\left[\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right] E\left[\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)\exp\left(-i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right] \\
&= E\left[\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)\left(y_t^+ - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\right)\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right] \\
&= -E\left[\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)\left(y_t^+ - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\right)\left(1 - \exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\right].
\end{aligned}$$

Then by the Fubini's theorem and Lemma 8.1, we obtain

$$\begin{aligned}
\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) &= -E\left[\begin{array}{c} \left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)\left(y_t^+ - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\right) \\ \times \int_{\mathbb{R}^q}\left(1 - \exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \end{array}\right] \\
&= -E\left[\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right)\left(y_t^+ - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right],
\end{aligned}$$

where $\left(y_t^+, \left(\mathbf{X}_t^+\right)'\right)'$ is an i.i.d. copy of $(y_t, \mathbf{X}_t')'$. ∎

The following Lemmas 8.2 to 8.6 further gives some important results regarding integrals involving the nonintegrable weighting function, which are useful in the proof of consistency and asymptotic normality of WCIV and WCIVF. Let

$$\begin{aligned}
\hat{\mathbf{Z}}_t\left(\boldsymbol{\tau}\right) &= \tilde{\mathbf{Y}}_t \exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right), \\
\mathbf{Z}_t\left(\boldsymbol{\tau}\right) &= \left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right),
\end{aligned}$$

where $\boldsymbol{\mu}_Y = E(\mathbf{Y}_t)$, $\tilde{\mathbf{Y}}_t = \mathbf{Y}_t - \hat{\boldsymbol{\mu}}_Y$, $\hat{\boldsymbol{\mu}}_Y = \frac{1}{n}\sum_{t=1}^n \mathbf{Y}_t$.

Lemma 8.2 *Let* $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{X}_t \in \mathbb{R}^q$. *If* $(\mathbf{Y}_t', \mathbf{X}_t')'$ *is i.i.d., and* $E\|\mathbf{Y}_t\|^2 < \infty$, $E\|\mathbf{X}_t\|^2 < \infty$. *Then*

$$\int_{\mathbb{R}^q} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\tau}) \xrightarrow{p} \int_{\mathbb{R}^q} E[\mathbf{Z}_t(\boldsymbol{\tau})]\,\boldsymbol{\omega}(d\boldsymbol{\tau}), \tag{21}$$

Proof. To prove (21), define the region $D(\delta) = \{\boldsymbol{\tau} : \delta \le \|\boldsymbol{\tau}\| \le 1/\delta\}$ with $\delta \in (0,1)$. For any fixed $\delta \in (0,1)$, $\omega(\boldsymbol{\tau})$ is bounded on $D(\delta)$. Hence by weak law of large number (WLLN) and the continuous mapping theorem it follows that

$$\int_{D(\delta)} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\omega(d\boldsymbol{\tau}) \xrightarrow{p} \int_{D(\delta)} E[\mathbf{Z}_t(\boldsymbol{\tau})]\,\omega(d\boldsymbol{\tau}).$$

It is obvious that $\int_{D(\delta)} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\omega(d\boldsymbol{\tau})$ converges in probability to $\int_{D(\delta)} E[\mathbf{Z}_t(\boldsymbol{\tau})]\,\omega(d\boldsymbol{\tau})$ when δ tends to zero.

Now it remains to show that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\left\|\int_{\mathbb{R}^q} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\tau}) - \int_{D(\delta)} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\omega(d\boldsymbol{\tau})\right\| = 0 \text{ in probability.}$$

For each $\delta \in (0,1)$, by triangle inequality,

$$\begin{aligned}
&\left\|\int_{\mathbb{R}^q} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\tau}) - \int_{D(\delta)} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\omega(d\boldsymbol{\tau})\right\| \\
&= \left\|\int_{\|\boldsymbol{\tau}\|<\delta} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\tau}) + \int_{\|\boldsymbol{\tau}\|>1/\delta} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\tau})\right\| \\
&\le \left\|\int_{\|\boldsymbol{\tau}\|<\delta} \frac{1}{n}\sum_{t=1}^n \tilde{\mathbf{Y}}_t\left[1 - \exp\left(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\right\| \\
&+ \left\|\int_{\|\boldsymbol{\tau}\|>1/\delta} \frac{1}{n}\sum_{t=1}^n \tilde{\mathbf{Y}}_t\left[1 - \exp\left(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\right\| \\
&:= A_{n1} + A_{n2}.
\end{aligned}$$

By triangle inequality,

$$
\begin{aligned}
A_{n1} &= \left\| \frac{1}{n} \sum_{t=1}^{n} \tilde{\mathbf{Y}}_t \int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right\| \\
&\leq \frac{1}{n} \sum_{t=1}^{n} \left\| \mathbf{Y}_t \int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right\| \\
&+ \left\| \frac{1}{n} \sum_{t=1}^{n} \mathbf{Y}_t \frac{1}{n} \sum_{t=1}^{n} \int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right\| \\
&\xrightarrow{p} E \left\| \mathbf{Y}_t \int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right\| \\
&+ \left\| E\mathbf{Y}_t E \int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right\|.
\end{aligned}
$$

Since $E\left(\int_{\mathbb{R}^q} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right)\right) = c_q E \left\|\mathbf{X}_t\right\| < \infty$, then

$$
\lim_{\delta \to 0} E \left(\int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right) = 0.
$$

By Cauchy-Schwarz inequality,

$$
E \left\| \mathbf{Y}_t \int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right\| \leq \left(E \left\|\mathbf{Y}_t\right\|^2\right)^{1/2} \left(E \left\| \int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right\|^2 \right)^{1/2}.
$$

Similarly, $E\left(\int_{\mathbb{R}^q} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right)\right)^2 = c_q^2 E \left\|\mathbf{X}_t\right\|^2 < \infty$,

$$
\lim_{\delta \to 0} E \left(\int_{\|\boldsymbol{\tau}\|<\delta} \left[1 - \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \right)^2 = 0.
$$

We have

$$
\lim_{\delta \to 0} \lim \sup_{n \to \infty} A_{n1} = 0 \text{ in probability.}
$$

Similarly, we have

$$
\lim_{\delta \to 0} \lim \sup_{n \to \infty} A_{n2} = 0 \text{ in probability.}
$$

We conclude that (21) holds. ■

In Lemma 8.3, the focus is on the process

$$
\mathbf{B}_{pn}\left(\boldsymbol{\tau}\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{\mathbf{Z}}_t\left(\boldsymbol{\tau}\right), \ \boldsymbol{\tau} \in \mathbb{R}^q.
$$

It is convenient to establish the weak convergence of $\mathbf{B}_{pn}(\boldsymbol{\tau})$ in a Hilbert space. By this approach the i.i.d. conditions can be relaxed to weakly stationary time series process conveniently. Specifically, for a fixed δ, $\omega(\cdot)$ is integrable on $D(\delta)$, therefore, denote υ as the product measure of $\omega(\cdot)$ on $D(\delta)$, i.e., $d\upsilon(\boldsymbol{\tau}) = \omega(d\boldsymbol{\tau})$ on $D(\delta)$. Then we consider $\mathbf{B}_{pn}(\boldsymbol{\tau})$ as a random element in the Hilbert space $L_2(D(\delta), \upsilon)$ of all square integrable q dimensional functions (with respect to the measure υ) with the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle_{H(\delta)} = \int_{D(\delta)} \mathbf{f}(\tau)' \mathbf{g}^c(\tau) \omega(d\boldsymbol{\tau}).$$

$L_2(D(\delta), \upsilon)$ is endowed with the natural Borel σ-field induced by the norm $\|\mathbf{f}\|_{H(\delta)} = \langle \mathbf{f}, \mathbf{f} \rangle_{H(\delta)}^{1/2}$. If $\mathbf{Z}$ is a $L_2(D(\delta), \upsilon)$-valued random element and has a probability ν_Z, we say $\mathbf{Z}$ has mean m and $E\left(\langle \mathbf{Z}, \mathbf{h} \rangle_{H(\delta)}\right) = \langle \mathbf{m}, \mathbf{h} \rangle_{H(\delta)}$ for any $\mathbf{h} \in L_2(D(\delta), \upsilon)$. If $E\|\mathbf{Z}\|_{H(\delta)}^2 < \infty$ and $\mathbf{Z}$ has zero mean, then the covariance operator of $\mathbf{Z}$ (or ν_Z), $\mathbf{C}_{\mathbf{Z}}(\cdot)$ say, is a continuous, linear, symmetric positive definite operator from $L_2(D(\delta), \upsilon)$ to $L_2(D(\delta), \upsilon)$, defined by

$$\mathbf{C}_{\mathbf{Z}}(h) = E\left[\langle \mathbf{Z}, \mathbf{h} \rangle_{H(\delta)} \mathbf{Z}\right].$$

An operator $\mathbf{s}$ on a Hilbert space is called nuclear if it can be represented as $\mathbf{s}(\mathbf{h}) = \sum_{j=1}^{\infty} l_j \langle \mathbf{h}, \mathbf{f}_j \rangle_{H(\delta)} \mathbf{f}_j$, where $\{\mathbf{f}_j\}$ is an orthonormal basis of the Hilbert space and $\{l_j\}$ is a real sequence, such that $\sum_{j=1}^{\infty} |l_j| < \infty$. It is easy to show, see, e.g., Bosq (2000), that the covariance operator $\mathbf{C}_{\mathbf{Z}}(\cdot)$ is a nuclear operator, provided that $E\|\mathbf{Z}\|_{H(\delta)}^2 < \infty$.

Lemma 8.3 *Let $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{X}_t \in \mathbb{R}^q$. If $(\mathbf{Y}_t', \mathbf{X}_t')'$ is i.i.d., $E(\mathbf{Y}_t|\mathbf{X}_t) = \boldsymbol{\mu}_Y$, and $E\|\mathbf{Y}_t\|^2 < \infty$, $E\|\mathbf{X}_t\|^2 < \infty$, then*

$$\mathbf{B}_{pn}(\boldsymbol{\tau}) \Rightarrow \mathbf{B}_p(\boldsymbol{\tau}), \tag{22}$$

where $\Rightarrow$ denotes weak convergence in $L_2(D(\delta), \upsilon)$, $\mathbf{B}_p(\cdot)$ denotes a zero-mean complex valued Gaussian process with a covariance structure given by

$$\begin{aligned}\Lambda_p(\boldsymbol{\tau}, \boldsymbol{\varsigma}) = {} & E\left[\mathbf{Y}_t\mathbf{Y}_t' \exp\left(i\langle \boldsymbol{\tau} - \boldsymbol{\varsigma}, \mathbf{X}_t\rangle\right)\right] - E(\mathbf{Y}_t) E\left(\mathbf{Y}_t'\right) E\left[\exp\left(i\langle \boldsymbol{\tau} - \boldsymbol{\varsigma}, \mathbf{X}_t\rangle\right)\right] \\ & + \left[E\left(\mathbf{Y}_t\mathbf{Y}_t'\right) + E(\mathbf{Y}_t) E\left(\mathbf{Y}_t'\right)\right] E\left[\exp\left(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle\right)\right] E\left[\exp\left(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle\right)\right] \\ & - E\left[\mathbf{Y}_t\mathbf{Y}_t' \exp\left(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle\right)\right] E\left[\exp\left(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle\right)\right] - E\left[\mathbf{Y}_t\mathbf{Y}_t' \exp\left(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle\right)\right] E\left[\exp\left(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle\right)\right],\end{aligned}$$

for $\boldsymbol{\tau}, \boldsymbol{\varsigma} \in D(\delta)$.

Proof. To prove (22), we show $\mathbf{B}_{pn}(\boldsymbol{\tau})$ is tight by Theorem 2.1 in Politis and Romano (1994). Firstly

$$E\left[\hat{\mathbf{Z}}_t(\boldsymbol{\tau})\right] = \frac{n-1}{n} E\left[\mathbf{Z}_t(\boldsymbol{\tau})\right] = 0.$$

For a fixed δ, by Cauchy-Schwarz inequality, the fact that $\|\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle)\|^2_{H(\delta)}$ is bounded, and $\|\mathbf{a}+\mathbf{b}\|^2_{H(\delta)} \le 2\|\mathbf{a}\|^2_{H(\delta)} + 2\|\mathbf{b}\|^2_{H(\delta)}$,

$$\begin{aligned}
E\left(\left\|\hat{\mathbf{Z}}_n(\boldsymbol{\tau})\right\|^2_{H(\delta)}\right) &\le E\left(\left\|\tilde{\mathbf{Y}}_t\right\|^2_{H(\delta)} \|\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle)\|^2_{H(\delta)}\right) \\
&\le CE\left(\left\|\mathbf{Y} - \frac{1}{n}\sum_{t=1}^{n}\mathbf{Y}_t\right\|^2_{H(\delta)}\right) \\
&\le 2CE\left(\|\mathbf{Y}_t\|^2_{H(\delta)} + \left\|\frac{1}{n}\sum_{t=1}^{n}\mathbf{Y}\right\|^2_{H(\delta)}\right) \\
&\le CE\|\mathbf{Y}_t\|^2 \le \infty.
\end{aligned}$$

For any integer $K > 1$, by WLLN, $\hat{\mathbf{Z}}_1(\boldsymbol{\tau}), ..., \hat{\mathbf{Z}}_K(\boldsymbol{\tau}) \xrightarrow{p} \mathbf{Z}_1(\boldsymbol{\tau}), ..., \mathbf{Z}_K(\boldsymbol{\tau})$.

$$\begin{aligned}
&\lim_{n\to\infty} E\left\langle \hat{\mathbf{Z}}_1(\boldsymbol{\tau}), \hat{\mathbf{Z}}_K(\boldsymbol{\tau})\right\rangle_{H(\delta)} \\
&= \lim_{n\to\infty} E\left\langle \mathbf{Z}_1(\boldsymbol{\tau}) - (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle), \mathbf{Z}_K(\boldsymbol{\tau}) - (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_K\rangle)\right\rangle_{H(\delta)} \\
&= E\left\langle \mathbf{Z}_1(\boldsymbol{\tau}), \mathbf{Z}_K(\boldsymbol{\tau})\right\rangle_{H(\delta)} - \lim_{n\to\infty} E\left\langle \mathbf{Z}_1(\boldsymbol{\tau}), (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_K\rangle)\right\rangle_{H(\delta)} \\
&- \lim_{n\to\infty} E\left\langle \mathbf{Z}_K(\boldsymbol{\tau}), (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_1\rangle)\right\rangle_{H(\delta)} \\
&+ \lim_{n\to\infty} E\left\langle (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_1\rangle), (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_K\rangle)\right\rangle_{H(\delta)} \\
&= \int_{D(\delta)} E\left[\mathbf{Z}_1(\boldsymbol{\tau})\right] E\left[\mathbf{Z}^c_K(\boldsymbol{\tau})\right]' \omega(d\boldsymbol{\tau}) \\
&= 0.
\end{aligned}$$

Since, for example,

$$
\begin{aligned}
& E\left\langle \mathbf{Z}_1(\boldsymbol{\tau}), (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_K\rangle)\right\rangle_{H(\delta)} \\
&= \int_{D(\delta)} E\left[\mathbf{Z}_1(\boldsymbol{\tau})'(\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_K\rangle)\right]\omega(d\boldsymbol{\tau}) \\
&= \int_{D(\delta)} E\left[\mathbf{Z}_1(\boldsymbol{\tau})'\frac{1}{n}\sum_{t=1}^{n}(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_K\rangle)\right]\omega(d\boldsymbol{\tau}) \\
&= \frac{1}{n}\int_{D(\delta)} E\left[\|\mathbf{Y}_1 - \boldsymbol{\mu}_Y\|^2\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_1 - \mathbf{X}_K\rangle)\right]\omega(d\boldsymbol{\tau}) \\
&\to 0, \text{ as } n \to \infty.
\end{aligned}
$$

Therefore

$$
\lim_{n\to\infty}\sum_{K=1}^{n} E\left\langle \hat{\mathbf{Z}}_1(\boldsymbol{\tau}), \hat{\mathbf{Z}}_K(\boldsymbol{\tau})\right\rangle_{H(\delta)} = E\left(\|\mathbf{Z}_1(\boldsymbol{\tau})\|^2_{H(\delta)}\right) < \infty.
$$

Further, for any $\mathbf{h} \in H(\delta)$,

$$
\begin{aligned}
\sigma^2_{n,h} &= Var\left(\langle \mathbf{B}_{pn}(\boldsymbol{\tau}), \mathbf{h}\rangle_{H(\delta)}\right) \\
&= \frac{1}{n}Var\left(\left\langle \sum_{t=1}^{n}\mathbf{Z}_t(\boldsymbol{\tau}) - (\hat{\boldsymbol{\mu}}_Y - \boldsymbol{\mu}_Y)\sum_{t=1}^{n}\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle), \mathbf{h}\right\rangle_{H(\delta)}\right) \\
&\to Var\left(\langle \mathbf{Z}_1(\boldsymbol{\tau}), \mathbf{h}\rangle_{H(\delta)}\right), \text{ as } n \to \infty.
\end{aligned}
$$

Then we conclude $\mathbf{B}_{pn}(\boldsymbol{\tau})$ is tight. Further, for any integer $K > 1$, $\mathbf{B}_{pn}(\boldsymbol{\tau}_1), ..., \mathbf{B}_{pn}(\boldsymbol{\tau}_K)$ are asymptotically normally distributed by the central limit theorem (CLT) and the Slutskey theorem. Then the weak convergence follows. Further

$$
\begin{aligned}
E\left[\mathbf{B}_{pn}(\boldsymbol{\tau})\mathbf{B}^c_{pn}(\boldsymbol{\varsigma})'\right] &= \left(\frac{n-1}{n}\right)^2 E\left[\mathbf{Y}_t\mathbf{Y}_t'\exp(i\langle \boldsymbol{\tau} - \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right] \\
&+ \frac{n-1}{n}E\left[\exp(i\langle \boldsymbol{\tau} - \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right]\left(\frac{1}{n}E\left(\mathbf{Y}_t\mathbf{Y}_t'\right) - E(\mathbf{Y}_t)E\left(\mathbf{Y}_t'\right)\right) \\
&+ \frac{n-1}{n}\left(E(\mathbf{Y}_t)E\left(\mathbf{Y}_t'\right) + \frac{n-2}{n}E\left(\mathbf{Y}_t\mathbf{Y}_t'\right)\right)E\left[\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right]E\left[\exp(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right] \\
&- \left(\frac{n-1}{n}\right)^2 E\left[\mathbf{Y}_t\mathbf{Y}_t'\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right]E\left[\exp(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right] \\
&- \left(\frac{n-1}{n}\right)^2 E\left[\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right]E\left[\mathbf{Y}_t\mathbf{Y}_t'\exp(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right].
\end{aligned}
$$

Then we have

$$\Lambda_p(\boldsymbol{\tau},\boldsymbol{\varsigma}) = E\left[\mathbf{Y}_t\mathbf{Y}_t'\exp\left(i\left\langle\boldsymbol{\tau}-\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right] - E\left(\mathbf{Y}_t\right)E\left(\mathbf{Y}_t'\right)E\left[\exp\left(i\left\langle\boldsymbol{\tau}-\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]$$
$$+\left[E\left(\mathbf{Y}_t\mathbf{Y}_t'\right)+E\left(\mathbf{Y}_t\right)E\left(\mathbf{Y}_t'\right)\right]E\left[\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]$$
$$-E\left[\mathbf{Y}_t\mathbf{Y}_t'\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]-E\left[\mathbf{Y}_t\mathbf{Y}_t'\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]$$

for $\boldsymbol{\tau},\boldsymbol{\varsigma}\in D(\delta)$. ■

Lemma 8.4 *Let $\mathbf{Y}_t\in\mathbb{R}^p$, $\mathbf{X}_t\in\mathbb{R}^q$. If $(\mathbf{Y}_t',\mathbf{X}_t')'$ is i.i.d., and $E\left\|\mathbf{Y}_t\right\|^2<\infty$, $E\left\|\mathbf{X}_t\right\|^2<\infty$. Then*

$$\int_{\mathbb{R}^q}\left\|E\left(\mathbf{Z}_t(\boldsymbol{\tau})\right)\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau}) = -E\left[\left(\mathbf{Y}_t-\boldsymbol{\mu}_Y\right)'\left(\mathbf{Y}_t^+-\boldsymbol{\mu}_Y\right)\left\|\mathbf{X}_t-\mathbf{X}_t^+\right\|\right],$$

$$\int_{\mathbb{R}^q}\left\|\frac{1}{n}\sum_{t=1}^n\hat{\mathbf{Z}}_t(\boldsymbol{\tau})\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau}) = -\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j'\tilde{\mathbf{Y}}_k\left\|X_j-X_k\right\|.$$

Further,

$$\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j'\tilde{\mathbf{Y}}_k\left\|X_j-X_k\right\|\xrightarrow{p}E\left[\left(\mathbf{Y}_t-\boldsymbol{\mu}_Y\right)'\left(\mathbf{Y}_t^+-\boldsymbol{\mu}_Y\right)\left\|\mathbf{X}_t-\mathbf{X}_t^+\right\|\right]. \tag{23}$$

if $E\left(\mathbf{Y}_t|\mathbf{X}_t\right)=\boldsymbol{\mu}_Y$, then

$$\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j'\tilde{\mathbf{Y}}_k\left\|X_j-X_k\right\| = O_p(1). \tag{24}$$

Proof. The analytical forms of $\int_{\mathbb{R}^q}\left\|E\left(\mathbf{Z}_t(\boldsymbol{\tau})\right)\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau})$ and $\int_{\mathbb{R}^q}\left\|\frac{1}{n}\sum_{t=1}^n\hat{\mathbf{Z}}_t(\boldsymbol{\tau})\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau})$ are proved by repeatedly applying Lemma 8.1. The proof of (23) follows the proof of Theorem 3 in Shao and Zhang (2014). To prove (24), we need to show

$$\int_{\mathbb{R}^q}\left\|\mathbf{B}_{np}(\boldsymbol{\tau})\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau})\xrightarrow{p}\int_{\mathbb{R}^q}\left\|\mathbf{B}_p(\boldsymbol{\tau})\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau}).$$

For a given δ, by Lemma 8.3 and the continuous mapping theorem, we have

$$\int_{D(\delta)}\left\|\mathbf{B}_{pn}(\boldsymbol{\tau})\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau})\xrightarrow{p}\int_{D(\delta)}\left\|\mathbf{B}_p(\boldsymbol{\tau})\right\|^2\boldsymbol{\omega}(d\boldsymbol{\tau}).$$

It is obvious that $\int_{D(\delta)}\left\|\mathbf{B}_{np}(\boldsymbol{\tau})\right\|^2\omega(d\boldsymbol{\tau})$ converges in distribution to $\int_{\mathbb{R}^q}\left\|\mathbf{B}_p(\boldsymbol{\tau})\right\|^2\omega(d\boldsymbol{\tau})$ when δ tends to zero.

For a given δ, following the proof of Theorem 4 in Shao and Zhang (2014), we have

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty} E\left|\int_{\|\boldsymbol{\tau}\|<\delta} \|\mathbf{B}_{pn}(\boldsymbol{\tau})\|^2 \omega(d\boldsymbol{\tau})\right| = 0,$$

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty} E\left|\int_{\|\boldsymbol{\tau}\|>1/\delta} \|\mathbf{B}_{pn}(\boldsymbol{\tau})\|^2 \omega(d\boldsymbol{\tau})\right| = 0.$$

Therefore

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty} E\left|\int_{\mathbb{R}^q} \|\mathbf{B}_{pn}(\boldsymbol{\tau})\|^2 \omega(d\boldsymbol{\tau}) - \int_{D(\delta)} \|\mathbf{B}_{pn}(\boldsymbol{\tau})\|^2 \omega(d\boldsymbol{\tau})\right| = 0.$$

Then by Markov's inequality,

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty} \left|\int_{\mathbb{R}^q} \|\mathbf{B}_{pn}(\boldsymbol{\tau})\|^2 \omega(d\boldsymbol{\tau}) - \int_{D(\delta)} \|\mathbf{B}_{pn}(\boldsymbol{\tau})\|^2 \omega(d\boldsymbol{\tau})\right| = 0 \text{ in probability.}$$

Finally, by Theorem 8.6.2 of Resnick (1999), we conclude that (24) holds ■

Lemma 8.5 *Let $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{X}_t \in \mathbb{R}^q$. If $(\mathbf{Y}_t', \mathbf{X}_t')'$ is i.i.d., and $E\|\mathbf{Y}_t\|^2 < \infty$, $E\|\mathbf{X}_t\|^2 < \infty$. Then*

$$\int_{\mathbb{R}^q} E[\mathbf{Z}_t(\boldsymbol{\tau})] E[\mathbf{Z}_t^c(\boldsymbol{\tau})]' \boldsymbol{\omega}(d\boldsymbol{\tau}) = -E\left[(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right],$$

$$\int_{\mathbb{R}^q} \frac{1}{n^2}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) \left(\sum_{t=1}^n \hat{\mathbf{Z}}_t^c(\boldsymbol{\tau})\right)' \boldsymbol{\omega}(d\boldsymbol{\tau}) = -\frac{1}{n^2}\sum_{j,k} \tilde{\mathbf{Y}}_j \tilde{\mathbf{Y}}_k' \|X_j - X_k\|.$$

Further,

$$\frac{1}{n^2}\sum_{j,k} \tilde{\mathbf{Y}}_j \tilde{\mathbf{Y}}_k' \|X_j - X_k\| \xrightarrow{p} E\left[(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right]. \tag{25}$$

If $E(\mathbf{Y}_t|\mathbf{X}_t) = \boldsymbol{\mu}_Y$, then

$$\int_{\mathbb{R}^q} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) \left(\sum_{t=1}^n \hat{\mathbf{Z}}_t^c(\boldsymbol{\tau})\right)' \boldsymbol{\omega}(d\boldsymbol{\tau}) = O_p(1). \tag{26}$$

Proof. The analytical forms are proved by repeatedly applying Lemma 8.1. The proof of (25) is analogous to the one for proving (23) in Lemma 8.4. To prove (26), we need to show

$$\int_{\mathbb{R}^q} \frac{1}{n}\sum_{t=1}^n \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) \left(\sum_{t=1}^n \hat{\mathbf{Z}}_t^c(\boldsymbol{\tau})\right)' \boldsymbol{\omega}(d\boldsymbol{\tau}) \xrightarrow{p} \int_{\mathbb{R}^q} \mathbf{B}_p(\boldsymbol{\tau}) \mathbf{B}_p^c(\boldsymbol{\tau})' \boldsymbol{\omega}(d\boldsymbol{\tau}).$$

Again, by Lemma 8.3 and the continuous mapping theorem, for a given $\delta \in (0,1)$, we have

$$\int_{D(\delta)} \frac{1}{n} \sum_{t=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) \left(\sum_{t=1}^{n} \hat{\mathbf{Z}}_t^c(\boldsymbol{\tau}) \right)' \boldsymbol{\omega}(d\boldsymbol{\tau}) \xrightarrow{p} \int_{D(\delta)} \mathbf{B}_p(\boldsymbol{\tau}) \mathbf{B}_p^c(\boldsymbol{\tau})' \boldsymbol{\omega}(d\boldsymbol{\tau}).$$

When $j = k$, from Lemma 8.4, we have

$$\lim_{\delta \to 0} \lim \sup_{n \to \infty} E\left(\int_{\|\boldsymbol{\tau}\| < \delta} \frac{1}{n} \left| \sum_{t=1}^{n} \hat{\mathbf{Z}}_{jt}(\boldsymbol{\tau}) \right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) \right) = 0, \text{for } j = 1, ..., p,$$

$$\lim_{\delta \to 0} \lim \sup_{n \to \infty} E\left(\int_{\|\boldsymbol{\tau}\| > 1/\delta} \frac{1}{n} \left| \sum_{t=1}^{n} \hat{\mathbf{Z}}_{jt}(\boldsymbol{\tau}) \right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) \right) = 0, \text{for } j = 1, ..., p,$$

where $\hat{Z}_{jt}(\boldsymbol{\tau})$ is the jth element of $\hat{\mathbf{Z}}_t(\boldsymbol{\tau})$. For $j, k = 1, ..., p$, $j \neq k$, by Cauchy-Schwarz inequality,

$$\begin{aligned}
& E\left| \int_{\|\boldsymbol{\tau}\| < \delta} \frac{1}{n} \sum_{t=1}^{n} \hat{Z}_{jt}(\boldsymbol{\tau}) \sum_{t=1}^{n} \hat{Z}_{kt}^c(\boldsymbol{\tau}) \boldsymbol{\omega}(d\boldsymbol{\tau}) \right| \\
& \leq E\left[\left(\int_{\|\boldsymbol{\tau}\| < \delta} \frac{1}{n} \left| \sum_{t=1}^{n} \hat{Z}_{jt}(\boldsymbol{\tau}) \right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) \right)^{1/2} \left(\int_{\|\boldsymbol{\tau}\| < \delta} \frac{1}{n} \left| \sum_{t=1}^{n} \hat{Z}_{kt}(\boldsymbol{\tau}) \right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) \right)^{1/2} \right] \\
& \leq \left[E\left(\int_{\|\boldsymbol{\tau}\| < \delta} \frac{1}{n} \left| \sum_{t=1}^{n} \hat{Z}_{jt}(\boldsymbol{\tau}) \right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) \right) \right]^{1/2} \left[E\left(\int_{\|\boldsymbol{\tau}\| < \delta} \frac{1}{n} \left| \sum_{t=1}^{n} \hat{Z}_{kt}(\boldsymbol{\tau}) \right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) \right) \right]^{1/2}.
\end{aligned}$$

So, by the dominated convergence theorem,

$$\lim_{\delta \to 0} \lim \sup_{n \to \infty} E\left| \int_{\|\boldsymbol{\tau}\| < \delta} \frac{1}{n} \sum_{t=1}^{n} \hat{Z}_{jt}(\boldsymbol{\tau}) \sum_{t=1}^{n} \hat{Z}_{kt}^c(\boldsymbol{\tau}) \boldsymbol{\omega}(d\boldsymbol{\tau}) \right| = 0.$$

Similarly we can obtain

$$\lim_{\delta \to 0} \lim \sup_{n \to \infty} E\left| \int_{\|\boldsymbol{\tau}\| > 1/\delta} \frac{1}{n} \sum_{t=1}^{n} \hat{Z}_{jt}(\boldsymbol{\tau}) \sum_{t=1}^{n} \hat{Z}_{kt}^c(\boldsymbol{\tau}) \boldsymbol{\omega}(d\boldsymbol{\tau}) \right| = 0.$$

Then for $j, k = 1, ..., p$,

$$\lim_{\delta \to 0} \lim \sup_{n \to \infty} E\left| \int_{\mathbb{R}^q} \frac{1}{n} \sum_{t=1}^{n} \hat{Z}_{jt}(\boldsymbol{\tau}) \sum_{t=1}^{n} \hat{Z}_{kt}^c(\boldsymbol{\tau}) \boldsymbol{\omega}(d\boldsymbol{\tau}) - \int_{D(\delta)} \frac{1}{n} \sum_{t=1}^{n} \hat{Z}_{jt}(\boldsymbol{\tau}) \sum_{t=1}^{n} \hat{Z}_{kt}^c(\boldsymbol{\tau}) \boldsymbol{\omega}(d\boldsymbol{\tau}) \right| = 0.$$

Then by Markov's inequality, $j,k=1,...,n$,

$$\lim_{\delta\to0}\limsup_{n\to\infty}\left(\int_{\mathbb{R}^q}\frac{1}{n}\sum_{t=1}^{n}\hat{Z}_{jt}(\boldsymbol{\tau})\sum_{t=1}^{n}\hat{Z}_{kt}^{c}(\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\tau})-\int_{D(\delta)}\frac{1}{n}\sum_{t=1}^{n}\hat{Z}_{jt}(\boldsymbol{\tau})\sum_{t=1}^{n}\hat{Z}_{kt}^{c}(\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\tau})\right)=0$$

in probability. Then by the continuous mapping theorem,

$$\lim_{\delta\to0}\limsup_{n\to\infty}\left(\int_{\mathbb{R}^q}\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbf{Z}}_{t}(\boldsymbol{\tau})\left(\sum_{t=1}^{n}\hat{\mathbf{Z}}_{t}^{c}(\boldsymbol{\tau})\right)'\boldsymbol{\omega}(d\boldsymbol{\tau})-\int_{D(\delta)}\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbf{Z}}_{t}(\boldsymbol{\tau})\left(\sum_{t=1}^{n}\hat{\mathbf{Z}}_{t}^{c}(\boldsymbol{\tau})\right)'\boldsymbol{\omega}(d\boldsymbol{\tau})\right)=0$$

in probability. Finally, by Theorem 8.6.2 of Resnick (1999), we conclude that (26) holds. ■

In the following Lemma, we give similar results without proving them.

Lemma 8.6 *Let $\mathbf{Y}_t\in\mathbb{R}^p$, $\mathbf{X}_t\in\mathbb{R}^q$. If $(\mathbf{Y}_t',\mathbf{X}_t')'$ is i.i.d., and $E\|\mathbf{Y}_t\|^2<\infty$, $E\|\mathbf{X}_t\|^2<\infty$, $E\|\mathbf{f}(\mathbf{X}_t)\|^2<\infty$. Let $\mathbf{F}_t(\boldsymbol{\tau})=(\mathbf{f}(\mathbf{X}_t)-E\mathbf{f}(\mathbf{X}_t))\exp(i\langle\boldsymbol{\tau},\mathbf{X}_t\rangle)$, $\hat{\mathbf{F}}_t(\boldsymbol{\tau})=\tilde{\mathbf{f}}(\mathbf{X}_t)\exp(-i\langle\boldsymbol{\tau},\mathbf{X}_t\rangle)$, where $\tilde{\mathbf{f}}(\mathbf{X}_k)=\mathbf{f}(\mathbf{X}_k)-\frac{1}{n}\sum_{j=1}^{n}\mathbf{f}(\mathbf{X}_j)$. Then*

$$\int_{\mathbb{R}^q}E[\mathbf{Z}_t(\boldsymbol{\tau})]\,E[\mathbf{F}_t^c(\boldsymbol{\tau})]'\boldsymbol{\omega}(d\boldsymbol{\tau})=-E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\left(\mathbf{f}\left(\mathbf{X}_t^+\right)-E\mathbf{f}(\mathbf{X}_t)\right)'\left\|\mathbf{X}_t-\mathbf{X}_t^+\right\|\right],$$

$$\int_{\mathbb{R}^q}\frac{1}{n^2}\sum_{t=1}^{n}\hat{\mathbf{Z}}_t(\boldsymbol{\tau})\sum_{t=1}^{n}\hat{\mathbf{F}}_t^c(\boldsymbol{\tau})'\boldsymbol{\omega}(d\boldsymbol{\tau})=-\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j\tilde{\mathbf{f}}(\mathbf{X}_k)'\|X_j-X_k\|.$$

Further,

$$\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j\tilde{\mathbf{f}}(\mathbf{X}_k)'\|X_j-X_k\|\xrightarrow{p}E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\left(\mathbf{f}\left(\mathbf{X}_t^+\right)-E\mathbf{f}(\mathbf{X}_t)\right)'\left\|\mathbf{X}_t-\mathbf{X}_t^+\right\|\right].$$

If $E(\mathbf{Y}_t|\mathbf{X}_t)=\boldsymbol{\mu}_Y$, then

$$\int_{\mathbb{R}^q}\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbf{Z}}_t(\boldsymbol{\tau})\sum_{t=1}^{n}\hat{\mathbf{F}}_t^c(\boldsymbol{\tau})'\boldsymbol{\omega}(d\boldsymbol{\tau})=O_p(1).\qquad(27)$$

The following lemma is Lemma A0 from Hansen et al. (2008).

Lemma 8.7 *If Assumptions 1 is satisfied, $\left\|\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)/r_n\right\|^2/\left(1+\left\|\hat{\boldsymbol{\beta}}\right\|^2\right)\xrightarrow{p}0$, then*

$$\left\|\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)/r_n\right\|\xrightarrow{p}0.$$

Lemma 8.8 *Let* $\tilde{\varepsilon}_j = \varepsilon_j - \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t$*, under Assumptions 1-3,*

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_j D_{jk}\tilde{\varepsilon}_k = o_p(1),$$

$$\frac{1}{n}\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime} = \frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jk}\tilde{\mathbf{f}}(\mathbf{X}_k)' + o_p(1).$$

Proof. Note $E(\varepsilon_t|\mathbf{X}_t) = 0$, $D_{jk} = -\|X_j - X_k\|$, then by Lemma 8.4,

$$\frac{1}{n}\sum_{j,k}\tilde{\varepsilon}_j D_{jk}\tilde{\varepsilon}_k = O_p(1).$$

Note $r_n = \min_{1\le j\le q} r_{j,n} \to \infty$, so we have

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_j D_{jk}\tilde{\varepsilon}_k = \frac{1}{r_n^2}O_p(1) = o_p(1).$$

$$\begin{aligned}\frac{1}{n}\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_j'\mathbf{R}_n^{-1\prime} &= \frac{1}{n}\sum_{j,k}\left(\frac{\tilde{\mathbf{f}}(\mathbf{X}_j)}{\sqrt{n}} + \mathbf{R}_n^{-1}\tilde{\boldsymbol{\eta}}_j\right)D_{jk}\left(\frac{\tilde{\mathbf{f}}(\mathbf{X}_k)'}{\sqrt{n}} + \tilde{\boldsymbol{\eta}}_k'\mathbf{R}_n^{-1\prime}\right)\\ &= \frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jk}\tilde{\mathbf{f}}(\mathbf{X}_k)' + \frac{1}{n}\sum_{j,k}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jk}\tilde{\boldsymbol{\eta}}_k'\frac{\mathbf{R}_n^{-1\prime}}{\sqrt{n}}\\ &+ \frac{\mathbf{R}_n^{-1}}{\sqrt{n}}\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\mathbf{f}}(\mathbf{X}_k)' + \mathbf{R}_n^{-1}\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\boldsymbol{\eta}}_k'\mathbf{R}_n^{-1\prime}.\end{aligned}$$

As $E(\boldsymbol{\eta}_t|\mathbf{X}_t) = 0$, then $\frac{1}{n}\sum_{j,k}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jk}\tilde{\boldsymbol{\eta}}_k' = O_p(1)$, $\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\mathbf{f}}(\mathbf{X}_k)' = O_p(1)$ by Lemma 8.6; $\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\boldsymbol{\eta}}_k' = O_p(1)$ by Lemma 8.5. Further $\mathbf{R}_n^{-1} = o_p(1)$ by Assumption 2. So

$$\begin{aligned}\frac{1}{n}\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime} &= \frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jk}\tilde{\mathbf{f}}(\mathbf{X}_k)' + O_p(1)\,o_p(1) + O_p(1)\,o_p(1) + O_p(1)\,o_p(1)\\ &= \frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jk}\tilde{\mathbf{f}}(\mathbf{X}_k)' + o_p(1).\end{aligned}$$

■

Lemma 8.9 *If Assumptions 1-3 are satisfied, then for* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$, $\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0$.

Proof. Following the same arguments as in the proof of Lemma A3 in Hausman et al. (2012), w.p.a.1 for all $\boldsymbol{\beta}$**,** we have

$$C \leq \frac{1}{n}\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right)'\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right) \leq C\left(1+\|\boldsymbol{\beta}\|^2\right).$$

On the other hand,

$$\begin{aligned}
\frac{1}{nr_n^2}\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right)'\mathbf{D}\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right) &= \frac{1}{nr_n^2}\sum_{j,k}\left(\tilde{y}_j-\tilde{\mathbf{Y}}_j'\boldsymbol{\beta}\right)' D_{jk}\left(\tilde{y}_k-\tilde{\mathbf{Y}}_k'\boldsymbol{\beta}\right) \\
&= \frac{1}{nr_n^2}\sum_{j,k}\left(\tilde{\mathbf{Y}}_j'\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)+\tilde{\varepsilon}_j\right)' D_{jk}\left(\tilde{\mathbf{Y}}_k'\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)+\tilde{\varepsilon}_k\right) \\
&= \frac{1}{nr_n^2}\left(\mathbf{R}_n'\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)\right)'\left(\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime}\right)\mathbf{R}_n'\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right) \\
&+ \frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_j D_{jk}\tilde{\varepsilon}_k + \left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)'\frac{2}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_k.
\end{aligned}$$

Note

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_k D_{jk}\tilde{\varepsilon}_j = \frac{1}{nr_n^2}\sum_{j,k}\frac{\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)}{\sqrt{n}}D_{jk}\tilde{\varepsilon}_k + \frac{1}{nr_n^2}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\varepsilon}_k.$$

Since $E\left(\left(\varepsilon_t,\boldsymbol{\eta}'\right)'|\mathbf{X}_t\right)=0$, by Lemma 8.5, $\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\varepsilon}_k = O_p(1)$, by Lemma 8.6, $\frac{1}{n}\sum_{j,k}\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)D_{jk}\tilde{\varepsilon}_k = O_p(1)$. Then we have

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_k D_{jk}\tilde{\varepsilon}_j = o_p(1).$$

By Lemma 8.8, $\frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_j D_{jk}\tilde{\varepsilon}_k = o_p(1)$. By Assumption 2, w.p.a.1,$\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)D_{jk}\tilde{\mathbf{f}}\left(\mathbf{X}_k\right)' \geq C\mathbf{I}_p$, we have, w.p.a.1,

$$\begin{aligned}
\frac{1}{nr_n^2}\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right)'\mathbf{D}\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right) &= \frac{1}{r_n^2}\left(\mathbf{R}_n'\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)\right)'\left(\frac{1}{n}\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime}\right)\mathbf{R}_n'\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)+o_p(1) \\
&\geq C\left\|\mathbf{R}_n'\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)/r_n\right\|^2.
\end{aligned}$$

Let

$$\hat{Q}(\boldsymbol{\beta}) = \frac{1}{r_n^2}\frac{\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right)'\mathbf{D}\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right)}{\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right)'\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\boldsymbol{\beta}\right)}.$$

Then by Lemma 8.8, and $\frac{1}{n}\sum_{j=1}^n\tilde{\varepsilon}_j^2 = O_p(1)$, we have

$$\left|\hat{Q}\left(\boldsymbol{\beta}_0\right)\right| = \left|\frac{\frac{1}{r_n^2 n}\sum_{j,k}\tilde{\varepsilon}_j D_{jk}\tilde{\varepsilon}_k}{\frac{1}{n}\sum_{j=1}^n\tilde{\varepsilon}_j^2}\right| \xrightarrow{p} 0. \quad (28)$$

Since $\hat{\boldsymbol{\beta}}_{WCIV} = \arg\min_{\beta} \hat{Q}(\boldsymbol{\beta})$, we have $\hat{Q}\left(\hat{\boldsymbol{\beta}}_{WCIV}\right) \leq \hat{Q}(\boldsymbol{\beta}_0)$. Therefore w.p.a.1

$$0 \leq \frac{\left\|\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}_{WCIV} - \boldsymbol{\beta}_0\right)/r_n\right\|^2}{1 + \left\|\hat{\boldsymbol{\beta}}_{WCIV}\right\|^2} \leq C\hat{Q}\left(\hat{\boldsymbol{\beta}}_{WCIV}\right) \leq C\hat{Q}(\boldsymbol{\beta}_0) \xrightarrow{p} 0,$$

implying

$$\frac{\left\|\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}_{WCIV} - \boldsymbol{\beta}_0\right)/r_n\right\|^2}{1 + \left\|\hat{\boldsymbol{\beta}}_{WCIV}\right\|^2} \xrightarrow{p} 0.$$

Then by Lemma 8.7, we arrive at the conclusion. ■

Lemma 8.10 *If Assumptions 1-3 are satisfied,* $\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0$, *then*

$$\frac{\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)'\mathbf{D}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)}{\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)'\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)} = o_p\left(r_n^2\right).$$

Proof. Firstly, by WLLN, we have

$$\frac{1}{n}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)'\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right) = O_p(1).$$

$$\begin{aligned}\frac{1}{nr_n^2}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)'\mathbf{D}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right) &= \frac{1}{nr_n^2}\sum_{j,k}\left(\tilde{y}_j - \tilde{\mathbf{Y}}_j'\hat{\boldsymbol{\beta}}\right)' D_{jk}\left(\tilde{y}_k - \tilde{\mathbf{Y}}_k'\hat{\boldsymbol{\beta}}\right) \\ &= \left(\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}\right)/r_n\right)'\left(\frac{1}{n}\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime}\right)\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}\right)/r_n \\ &+ \frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_j D_{jk}\tilde{\varepsilon}_k + \left(r_n\mathbf{R}_n'^{-1}\right)\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}\right)/r_n\frac{2}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_k.\end{aligned}$$

By Lemma 8.8, $\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0$, $\left\|\mathbf{R}_n^{-1}\right\| = O\left(r_n^{-1}\right)$, and $\frac{1}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_k D_{jk}\tilde{\varepsilon}_j = o_p(1)$, we have

$$\frac{1}{n}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)'\mathbf{D}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right) = o_p\left(r_n^2\right).$$

Then by the continuous mapping theorem, the result follows. ■

Proof of Theorem 4.1. Note firstly when $\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0$, then by $\vartheta_{\min}\left(\mathbf{R}_n\mathbf{R}_n'/r_n^2\right) \geq$

$\vartheta_{\min}\left(\tilde{\mathbf{R}}_n\tilde{\mathbf{R}}_n'\right) > 0$, we have

$$\left\|\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)/r_n\right\| \geq \vartheta_{\min}\left(\mathbf{R}_n\mathbf{R}_n'/r_n^2\right)\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\| \geq C\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|,$$

implying $\hat{\boldsymbol{\beta}} \xrightarrow{\mathbf{p}} \boldsymbol{\beta}_0$. Therefore, for WCIV, this follows from Lemma 8.9. For WCIVF, note that firstly

$$\hat{\lambda}_{WCIV} = \frac{\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}_{WCIV}\right)'\mathbf{D}\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}_{WCIV}\right)}{\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}_{WCIV}\right)'\left(\tilde{\mathbf{y}}-\tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}_{WCIV}\right)} = o_p\left(r_n^2\right).$$

Then

$$\hat{\lambda}_{WCIVF} = \left[\hat{\lambda}_{WCIV} - \left(1-\hat{\lambda}_{WCIV}\right)C/n\right] / \left[1-\left(1-\hat{\lambda}_{WCIV}\right)C/n\right] = o_p\left(r_n^2\right).$$

$$\begin{aligned}
&\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}_{WCIVF}-\boldsymbol{\beta}_0\right)/r_n \\
&= \mathbf{R}_n'\left[\tilde{\mathbf{Y}}'\left(\mathbf{D}-\hat{\lambda}_{WCIVF}\mathbf{I}_n\right)\tilde{\mathbf{Y}}\right]^{-1}\left[\tilde{\mathbf{Y}}'\left(\mathbf{D}-\hat{\lambda}_{WCIVF}\mathbf{I}_n\right)\tilde{\mathbf{y}}\right]/r_n \\
&= \left[\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k' - \frac{1}{n}\hat{\lambda}_{WCIVF}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\right)\mathbf{R}_n^{-1\prime}\right]^{-1} \\
&\times \mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_k - \frac{1}{n}\hat{\lambda}_{WCIVF}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}\right)/r_n.
\end{aligned}$$

Since $\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} = O_p(n)$, $\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}} = O_p(n)$, $\left\|\mathbf{R}_n^{-1}\right\| = O\left(r_n^{-1}\right)$, therefore

$$\frac{\mathbf{R}_n^{-1}}{n}\hat{\lambda}_{WCIVF}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}R_n^{-1\prime} = O\left(r_n^{-1}\right)O\left(1/n\right)o_p\left(r_n^2\right)O_p\left(n\right)O\left(r_n^{-1}\right) = o_p\left(1\right),$$

$$\mathbf{R}_n^{-1}\frac{1}{n}\hat{\lambda}_{WCIVF}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}/r_n = O\left(r_n^{-1}\right)O\left(1/n\right)o_p\left(r_n^2\right)O_p\left(n\right)O\left(r_n^{-1}\right) = o_p\left(1\right).$$

$$\begin{aligned}
\mathbf{R}_n^{-1}\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_k/r_n &= \frac{1}{n}\sum_{j,k}\frac{\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)}{\sqrt{n}}D_{jk}\tilde{\varepsilon}_k/r_n + \mathbf{R}_n^{-1}\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\varepsilon}_k/r_n \\
&= o_p\left(1\right) + o_p\left(1\right) = o_p\left(1\right).
\end{aligned}$$

Further, by Lemma 8.8, we have

$$\mathbf{R}_n' \left(\hat{\boldsymbol{\beta}}_{WCIVF} - \boldsymbol{\beta}_0\right)/r_n = \left(\frac{1}{n^2}\sum_{j,k} \tilde{\mathbf{f}}\left(\mathbf{X}_j\right) D_{jk} \tilde{\mathbf{f}}\left(\mathbf{X}_k\right)' + o_p\left(1\right)\right)^{-1} o_p\left(1\right) = o_p\left(1\right).$$

Therefore, $\hat{\boldsymbol{\beta}}_{WCIVF} \xrightarrow{p} \boldsymbol{\beta}_0$. Finally, by the continuous mapping theorem, $\hat{\alpha}_{WCIV} \xrightarrow{p} \alpha_0$, $\hat{\alpha}_{WCIVF} \xrightarrow{p} \alpha_0$. ■

Proof of Theorem 4.2. Note, for $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ or $\hat{\boldsymbol{\beta}}_{WCIVF}$, $\hat{\lambda} = \hat{\lambda}_{WCIV}$ or $\hat{\lambda}_{WCIVF}$

$$\hat{\boldsymbol{\beta}} = \left[\tilde{\mathbf{Y}}'\left(\mathbf{D} - \hat{\lambda}\mathbf{I}_n\right)\tilde{\mathbf{Y}}\right]^{-1}\left[\tilde{\mathbf{Y}}'\left(\mathbf{D} - \hat{\lambda}\mathbf{I}_n\right)\tilde{\mathbf{y}}\right].$$

Then

$$\begin{aligned}
&\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \\
&= \mathbf{R}_n'\left[\tilde{\mathbf{Y}}'\left(\mathbf{D} - \hat{\lambda}\mathbf{I}_n\right)\tilde{\mathbf{Y}}\right]^{-1}\left[\tilde{\mathbf{Y}}'\left(\mathbf{D} - \hat{\lambda}\mathbf{I}_n\right)\tilde{\boldsymbol{\varepsilon}}\right] \\
&= \left[\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k' - \frac{1}{n}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\right)\mathbf{R}_n^{-1\prime}\right]^{-1} \\
&\times \mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_k - \frac{1}{n}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}\right).
\end{aligned}$$

$$\frac{\mathbf{R}_n^{-1}}{n}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\mathbf{R}_n^{-1\prime} = O\left(r_n^{-1}\right)O\left(1/n\right)o_p\left(r_n^2\right)O_p\left(n\right)O\left(r_n^{-1}\right) = o_p\left(1\right).$$

$$\begin{aligned}
\mathbf{R}_n^{-1}\frac{1}{\sqrt{n}}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}} &= \mathbf{R}_n^{-1}\frac{1}{n}\hat{\lambda}\sum_j \tilde{\mathbf{Y}}_j\tilde{\varepsilon}_j \\
&= \frac{1}{n}\hat{\lambda}\frac{1}{\sqrt{n}}\sum_j \tilde{\mathbf{f}}\left(\mathbf{X}_j\right)\tilde{\varepsilon}_k + \frac{1}{n}\hat{\lambda}\mathbf{R}_n^{-1}\sum_j \tilde{\boldsymbol{\eta}}_j\tilde{\varepsilon}_j \\
&= O\left(1/n\right)o_p\left(r_n^2\right)O_p\left(1\right) + O\left(1/\sqrt{n}\right)o_p\left(r_n^2\right)O\left(r_n^{-1}\right)O_p\left(1\right) \\
&= o_p\left(1\right),
\end{aligned}$$

Since by CLT, $\frac{1}{\sqrt{n}}\sum_j \tilde{\boldsymbol{\eta}}_j\tilde{\varepsilon}_j = O_p\left(1\right)$, $\frac{1}{\sqrt{n}}\sum_j \tilde{\mathbf{f}}\left(\mathbf{X}_j\right)\tilde{\varepsilon}_k = O_p\left(1\right)$, and $r_{j,n} = \sqrt{n}$ or $r_{j,n}/\sqrt{n} \to 0$

by Assumption 2. Therefore

$$\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)=\left[\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\right)\mathbf{R}_n^{-1\prime}\right]^{-1}\mathbf{R}_n^{-1}\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_k+o_p(1).$$

By Lemma 8.8, we have

$$\begin{aligned}&\left[\sqrt{n}\mathbf{R}_n^{-1}\left(\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\right)\mathbf{R}_n^{-1\prime}\sqrt{n}\right]\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\\&=\sqrt{n}\mathbf{R}_n^{-1}\frac{1}{n\sqrt{n}}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_k+o_p(1).\\&:=\mathbf{A}_n+o_p(1).\end{aligned}$$

Now

$$\begin{aligned}\mathbf{A}_n&=\sqrt{n}\mathbf{R}_n^{-1}\int_{\mathbb{R}^q}\frac{1}{n\sqrt{n}}\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_j\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_j\right\rangle\right)\tilde{\varepsilon}_k\exp\left(-i\left\langle\boldsymbol{\tau},\mathbf{X}_k\right\rangle\right)\boldsymbol{\omega}(d\boldsymbol{\tau})\\&=\sqrt{n}\mathbf{R}_n^{-1}\int_{\mathbb{R}^q}\frac{1}{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_j\right\rangle\right)\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\tilde{\varepsilon}_k\exp\left(-i\left\langle\boldsymbol{\tau},\mathbf{X}_k\right\rangle\right)\boldsymbol{\omega}(d\boldsymbol{\tau})\\&=\sqrt{n}\mathbf{R}_n^{-1}\int_{\mathbb{R}^q}\frac{1}{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_j\right\rangle\right)B_{1n}(\boldsymbol{\tau})^c\,\boldsymbol{\omega}(d\boldsymbol{\tau}),\end{aligned}$$

where

$$B_{1n}(\boldsymbol{\tau})=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\tilde{\varepsilon}_k\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_k\right\rangle\right).$$

By Lemma 8.4

$$B_{1n}(\boldsymbol{\tau})\Rightarrow B_1(\boldsymbol{\tau}),$$

where $B_1(\cdot)$ denotes a zero-mean complex valued Gaussian process with a covariance structure given by

$$\begin{aligned}\Lambda_1(\boldsymbol{\tau},\boldsymbol{\varsigma})&=E\left[\varepsilon_t^2\exp\left(i\left\langle\boldsymbol{\tau}-\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]+E\left(\varepsilon_t^2\right)E\left[\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]\\&\quad-E\left[\varepsilon_t^2\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]-E\left[\varepsilon_t^2\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right],\end{aligned}$$

for $\boldsymbol{\tau},\boldsymbol{\varsigma}\in D(\delta)$.

For a fixed δ, by the Slutsky theorem and the continuous mapping theorem,

$$\int_{D(\delta)} \frac{1}{n} \sum_{j=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) B_{1n}(\boldsymbol{\tau})^c \boldsymbol{\omega}(d\boldsymbol{\tau}) \xrightarrow{p} \int_{D(\delta)} E[\mathbf{Z}_t(\boldsymbol{\tau})] B_1(\boldsymbol{\tau})^c \boldsymbol{\omega}(d\boldsymbol{\tau}).$$

Then in the same spirit of proving 8.5, by Cauchy-Schwarz inequality

$$\begin{aligned}
& E\left|\int_{\|\boldsymbol{\tau}\|<\delta} \frac{1}{n} \sum_{t=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) B_{1n}(\boldsymbol{\tau})^c \boldsymbol{\omega}(d\boldsymbol{\tau})\right| \\
& \leq E\left[\left(\int_{\|\boldsymbol{\tau}\|<\delta} \frac{1}{n^2}\left|\sum_{t=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau})\right)^{1/2}\left(\int_{\|\boldsymbol{\tau}\|<\delta}\left|B_{1n}(\boldsymbol{\tau})\right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau})\right)^{1/2}\right] \\
& \leq\left(E\left(\int_{\|\boldsymbol{\tau}\|<\delta} \frac{1}{n^2}\left|\sum_{t=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau})\right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau})\right)\right)^{1/2}\left(E\left(\int_{\|\boldsymbol{\tau}\|<\delta}\left|B_{1n}(\boldsymbol{\tau})\right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau})\right)\right)^{1/2}.
\end{aligned}$$

Then by the dominated convergence theorem,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E\left|\int_{\|\boldsymbol{\tau}\|<\delta} \frac{1}{n} \sum_{t=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) B_{1n}(\boldsymbol{\tau})^c \boldsymbol{\omega}(d\boldsymbol{\tau})\right| = 0.$$

Similarly we can obtain

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E\left|\int_{\|\boldsymbol{\tau}\|>1/\delta} \frac{1}{n} \sum_{t=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) B_{1n}(\boldsymbol{\tau})^c \boldsymbol{\omega}(d\boldsymbol{\tau})\right| = 0.$$

So we conclude that

$$\frac{\mathbf{R}_n}{\sqrt{n}} \mathbf{A}_n \xrightarrow{d} \int_{\mathbb{R}^q} E[\mathbf{Z}_t(\boldsymbol{\tau})] B_1(\boldsymbol{\tau})^c \boldsymbol{\omega}(d\boldsymbol{\tau}).$$

Given the fact that the integrated weighted Gaussian process follows a normal distribution, so we have

$$\frac{\mathbf{R}_n}{\sqrt{n}} Var(\mathbf{A}_n) \frac{\mathbf{R}_n'}{\sqrt{n}} \xrightarrow{p} \boldsymbol{\Omega}(\boldsymbol{\theta}_0)$$

where

$$\begin{aligned}
\boldsymbol{\Omega}(\boldsymbol{\theta}_0) &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} E[\mathbf{Z}_t(\boldsymbol{\tau})] E[\mathbf{Z}_t(-\boldsymbol{\varsigma})]' \Lambda_1(\boldsymbol{\tau}, \boldsymbol{\varsigma})^c \boldsymbol{\omega}(d\boldsymbol{\tau}) \boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= \mathbf{S}_1(\boldsymbol{\theta}_0) + \mathbf{S}_2(\boldsymbol{\theta}_0) - \mathbf{S}_3(\boldsymbol{\theta}_0) - \mathbf{S}_3(\boldsymbol{\theta}_0)'.
\end{aligned}$$

To derive the analytical form of $\mathbf{\Omega}(\boldsymbol{\theta}_0)$, we plug $\Lambda_1(\boldsymbol{\tau},\boldsymbol{\varsigma})^c$ into $\mathbf{\Omega}(\boldsymbol{\theta}_0)$:

$$\mathbf{\Omega}(\boldsymbol{\theta}_0) = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E[\mathbf{Z}_t(\boldsymbol{\tau})]\, E[\mathbf{Z}_t(-\boldsymbol{\varsigma})]' \times \left\{ \begin{array}{c} E\left[\varepsilon_t^2 \exp(-i\langle \boldsymbol{\tau}-\boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right] + E\left(\varepsilon_t^2\right) E[\exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)]\, E[\exp(i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)] \\ -E\left[\varepsilon_t^2 \exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right] E[\exp(i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)] - E\left[\varepsilon_t^2 \exp(i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right] E[\exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)] \end{array} \right\} \boldsymbol{\omega}(d\boldsymbol{\tau})\, \boldsymbol{\omega}(d\boldsymbol{\varsigma}),$$

By the Fubini's theorem and Lemma 8.1,

$$\begin{aligned}
\mathbf{S}_1(\boldsymbol{\theta}_0) &= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\begin{array}{c} (\mathbf{Y}_t - \boldsymbol{\mu}_Y)\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^+\right\rangle\right) \\ \times \varepsilon_t^{++2}\exp\left(-i\left\langle \boldsymbol{\tau}-\boldsymbol{\varsigma}, \mathbf{X}_t^{++}\right\rangle\right) \end{array} \right] \boldsymbol{\omega}(d\boldsymbol{\tau})\, \boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\begin{array}{c} \varepsilon_t^{++2}(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^{++}\right\rangle\right) \\ \times \left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)'\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^{++} - \mathbf{X}_t^+\right\rangle\right) \end{array} \right] \boldsymbol{\omega}(d\boldsymbol{\tau})\, \boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\begin{array}{c} \varepsilon_t^{++2}(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left(1-\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^{++}\right\rangle\right)\right) \\ \times \left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)'\left(1-\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^{++} - \mathbf{X}_t^+\right\rangle\right)\right) \end{array} \right] \boldsymbol{\omega}(d\boldsymbol{\tau})\, \boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= E\left[\begin{array}{c} \int_{\mathbb{R}^q} \varepsilon_t^{++2}(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left[1-\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^{++}\right\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau}) \\ \times \int_{\mathbb{R}^q}\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)'\left[1-\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^{++} - \mathbf{X}_t^+\right\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma}) \end{array} \right] \\
&= E\left[\varepsilon_t^{++2}(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)' \left\|\mathbf{X}_t - \mathbf{X}_t^{++}\right\| \left\|\mathbf{X}_t^+ - \mathbf{X}_t^{++}\right\|\right].
\end{aligned}$$

$$\begin{aligned}
\mathbf{S}_2(\boldsymbol{\theta}_0) &= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E[(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)]\, E[(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\exp(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)]' \\
&\quad \times E\left(\varepsilon_t^2\right) E[\exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)]\, E[\exp(i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)]\, \boldsymbol{\omega}(d\boldsymbol{\tau})\, \boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= E\left(\varepsilon_t^2\right)\int_{\mathbb{R}^q} E\left[(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left(1-\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau}) \\
&\quad \times \int_{\mathbb{R}^q} E\left[(\mathbf{Y}_t - \boldsymbol{\mu}_Y)'\left(1-\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^+ - \mathbf{X}_t\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= E\left(\varepsilon_t^2\right) E\left((\mathbf{Y}_t - \boldsymbol{\mu}_Y)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right) E\left((\mathbf{Y}_t - \boldsymbol{\mu}_Y)'\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right).
\end{aligned}$$

$$
\begin{aligned}
\mathbf{S}_3(\boldsymbol{\theta}_0) &= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau},\mathbf{X}_t\rangle)\right] E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\exp(-i\langle\boldsymbol{\varsigma},\mathbf{X}_t\rangle)\right]' \\
&\times E\left[\varepsilon_t^2\exp(-i\langle\boldsymbol{\tau},\mathbf{X}_t\rangle)\right] E\left[\exp(i\langle\boldsymbol{\varsigma},\mathbf{X}_t\rangle)\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= \int_{\mathbb{R}^q} E\left[\varepsilon_t^{+2}(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\left(1-\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t-\mathbf{X}_t^+\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau}) \\
&\times \int_{\mathbb{R}^q} E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)'\left(1-\exp\left(i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t^+-\mathbf{X}_t\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= E\left(\varepsilon_t^{+2}(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\left\|\mathbf{X}_t-\mathbf{X}_t^+\right\|\right) E\left((\mathbf{Y}_t-\boldsymbol{\mu}_Y)'\left\|\mathbf{X}_t-\mathbf{X}_t^+\right\|\right).
\end{aligned}
$$

The last term in $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$:

$$
\begin{aligned}
&\int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\exp(i\langle\boldsymbol{\tau},\mathbf{X}_t\rangle)\right] E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\exp(-i\langle\boldsymbol{\varsigma},\mathbf{X}_t\rangle)\right]' \\
&\times E\left[\varepsilon_t^2\exp(i\langle\boldsymbol{\varsigma},\mathbf{X}_t\rangle)\right] E\left[\exp(-i\langle\boldsymbol{\tau},\mathbf{X}_t\rangle)\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= \int_{\mathbb{R}^q} E\left[(\mathbf{Y}_t-\boldsymbol{\mu}_Y)\left(1-\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t-\mathbf{X}_t^+\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau}) \\
&\times \int_{\mathbb{R}^q} E\left[\varepsilon_t^{+2}(\mathbf{Y}_t-\boldsymbol{\mu}_Y)'\left(1-\exp\left(i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t^+-\mathbf{X}_t\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma}) \\
&= \mathbf{S}_3(\boldsymbol{\theta}_0)'.
\end{aligned}
$$

Finally by the Slutsky theorem, we arrive at the conclusion. ■

Proof of Theorem 4.3. For $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_{WCIV}$ or $\hat{\boldsymbol{\theta}}_{WCIVF}$, $\hat{\lambda}$=$\hat{\lambda}_{WCIV}$ or $\hat{\lambda}_{WCIVF}$,

$$
\begin{aligned}
\hat{\mathbf{S}}_1\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right) &= \frac{1}{n^3}\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n \tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2\tilde{D}_{jl}\left(\hat{\lambda}\right)\tilde{D}_{kl}\left(\hat{\lambda}\right) \\
&= \frac{1}{n^3}\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n \tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 D_{jl}D_{kl} \\
&- \frac{1}{n^3}\hat{\lambda}\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n \tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 D_{jl}I_{kl} \\
&- \frac{1}{n^3}\hat{\lambda}\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n \tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 D_{kl}I_{jl} \\
&+ \frac{1}{n^3}\hat{\lambda}^2\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n \tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 I_{kl}I_{jl} \\
&:= \mathbf{A}_{1n}-\mathbf{A}_{2n}-\mathbf{A}_{3n}+\mathbf{A}_{4n}.
\end{aligned}
$$

where I_{kl} denotes the (k, l)th element of $\mathbf{I}_n$.

$$\begin{aligned}
A_{2n} &= \frac{1}{n^3}\hat{\lambda}\sum_{l=1}^{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_l'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 D_{jl} \\
&= \frac{\hat{\lambda}}{n}\frac{1}{n^2}\sum_{l=1}^{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_l'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 D_{jl} \\
&= o_p(1)\,O_p(1) \\
&= o_p(1),
\end{aligned}$$

since by Lemma 8.10, $\hat{\lambda} = o_p\left(r_n^2\right)$, and

$$\begin{aligned}
\frac{1}{n^2}\sum_{l=1}^{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_l'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 D_{jl} &= \frac{1}{n^2}\sum_{l=1}^{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_j'\varepsilon_l^2 D_{jl} + o_p(1) \\
&= E\left[\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)'\varepsilon_t^{+2}\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right] + o_p(1) \\
&= O_p(1).
\end{aligned}$$

By similar arguments, we have $\mathbf{A}_{3n} = o_p(1)$.

$$\begin{aligned}
\mathbf{A}_{4n} &= \frac{1}{n^3}\hat{\lambda}^2\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 I_{kl}I_{jl} \\
&= \frac{1}{n^3}\hat{\lambda}^2\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_j'\varepsilon_j\left(\hat{\boldsymbol{\theta}}\right)^2 \\
&= o_p(1).
\end{aligned}$$

Now

$$\mathbf{A}_{1n} = \frac{1}{n}\sum_{l=1}^{n}\left(\varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2\frac{1}{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j D_{jl}\frac{1}{n}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_k' D_{kl}\right).$$

We have

$$\frac{1}{n}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_j D_{jl} \xrightarrow{p} \int_{\mathbb{R}^q} E_j\left[\left(\mathbf{Y}_j - \boldsymbol{\mu}_Y\right)\exp\left(i\left\langle\boldsymbol{\tau}, \mathbf{X}_j - \mathbf{X}_l\right\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau}) = -E_j\left[\left(\mathbf{Y}_j - \boldsymbol{\mu}_Y\right)\left\|\mathbf{X}_j - \mathbf{X}_l\right\|\right],$$

$$\frac{1}{n}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_k' D_{kl} \xrightarrow{p} \int_{\mathbb{R}^q} E_k\left[\left(\mathbf{Y}_k - \boldsymbol{\mu}_Y\right)'\exp\left(i\left\langle\boldsymbol{\varsigma}, \mathbf{X}_k - \mathbf{X}_l\right\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma}) = -E_k\left[\left(\mathbf{Y}_k - \boldsymbol{\mu}_Y\right)'\left\|\mathbf{X}_k - \mathbf{X}_l\right\|\right],$$

where E_j denotes the expectation in terms of $(\mathbf{Y}_j, \mathbf{X}_j)$. So by the continuous mapping theorem, we conclude that $\hat{\mathbf{S}}_1\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\xrightarrow{p}\mathbf{S}_1\left(\boldsymbol{\theta}_0\right)$. Analogously we can show

$$\hat{\mathbf{S}}_2\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\xrightarrow{p}\mathbf{S}_2\left(\boldsymbol{\theta}_0\right),$$

$$\hat{\mathbf{S}}_3\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\xrightarrow{p}\mathbf{S}_3\left(\boldsymbol{\theta}_0\right).$$

Then by the continuous mapping theorem,

$$\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\xrightarrow{p}\boldsymbol{\Omega}\left(\boldsymbol{\theta}_0\right).$$

$$\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)=\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'-\frac{1}{n^2}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}.$$

Since $\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'=\Upsilon+o_p\left(1\right)$ by Lemma 8.5, $\frac{1}{n^2}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}=O\left(1/n^2\right)o_p\left(r_n^2\right)O\left(n\right)=o_p\left(1\right)$, we have

$$\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\xrightarrow{p}\Upsilon.$$

By the Slutsky theorem

$$\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\sqrt{n}\mathbf{R}_n^{-1}\right)^{-1/2}\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1}\sqrt{n}\right)\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\xrightarrow{d}N\left(0,\mathbf{I}_p\right).$$

In other words,

$$\mathbf{R}_n^{-1}\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1}\sqrt{n}\right)^{-1}\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\sqrt{n}\mathbf{R}_n^{-1}\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1}\sqrt{n}\right)^{-1}\mathbf{R}_n^{-1\prime}$$

is a consistent variance estimator for $\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)$. On the other hand, by the first-order Taylor expansion, under H_0,

$$\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)=\mathbf{g}\left(\boldsymbol{\beta}_0\right)+\mathbf{G}\left(\bar{\boldsymbol{\beta}}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)=\mathbf{G}\left(\bar{\boldsymbol{\beta}}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right),$$

where $\bar{\boldsymbol{\beta}}$ is vector between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$, $\bar{\boldsymbol{\beta}}\xrightarrow{p}\boldsymbol{\beta}_0$. Then

$$\begin{aligned}
&\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{R}_n^{-1}\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1}\sqrt{n}\right)^{-1}\\
&\times\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\sqrt{n}\mathbf{R}_n^{-1}\left(\sqrt{n}\mathbf{R}_n^{-1\prime}\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1}\sqrt{n}\right)^{-1}\mathbf{R}_n^{-1\prime}\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)'\\
&=\frac{1}{n}\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)'
\end{aligned}$$

is a consistent variance estimator of $\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)$. Therefore

$$W_n\left(\hat{\boldsymbol{\theta}}\right) \xrightarrow{d} \chi^2_m.$$

■

References

Ackerberg, D. A. and Devereux, P. J. (2009). Improved jive estimators for overidentified linear models with and without heteroskedasticity. *The Review of Economics and Statistics*, 91(2):351–362.

Ai, H. (2010). Information quality and long-run risk: Asset pricing implications. *Journal of Finance*, 65(4):1333–1367.

Anatolyev, S. (2019). Many instruments and/or regressors: A friendly guide. *Journal of Economic Surveys*, 33(2):689–726.

Anderson, T., Kunitomo, N., and Matsushita, Y. (2010). On the asymptotic optimality of the liml estimator with possibly many instruments. *Journal of Econometrics*, 157(2):191–204.

Angrist, J. D., Imbens, G. W., and Krueger, A. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67.

Antoine, B. and Lavergne, P. (2014). Conditional moment models under semi-strong identification. *Journal of Econometrics*, 182(1):59–69.

Ascari, G., Magnusson, L. M., and Mavroeidis, S. (2021). Empirical evidence on the euler equation for consumption in the us. *Journal of Monetary Economics*, 117:129–152.

Attanasio, O. P. and Weber, G. (1993). Consumption growth, the interest rate and aggregation. *The Review of Economic Studies*, 60(3):631–649.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59(4):1481–1509.

Barro, R. J. (2009). Rare disasters, asset prices, and welfare costs. *The American Economic Review*, 99(1):243–264.

Beeler, J. and Campbell, J. Y. (2012). The long-run risks model and aggregate asset prices: An empirical assessment. *Critical Finance Review*, 1(1):141–182.

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

Bekker, P. A. and van der Ploeg, J. (2005). Instrumental variable estimation based on grouped data. *Statistica Neerlandica*, 59(3):239–267.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58(6):1443–1455.

Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65(5):1129–1151.

Blomquist, S. and Dahlberg, M. (1999). Small sample properties of liml and jackknife iv estimators: Experiments with weak instruments. *Journal of Applied Econometrics*, 14(1):69–88.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Springer-Verlag, New York.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

Campbell, J. Y. (2003). Consumption-based asset pricing. *Handbook of The Economics of Finance*, pages 803–887.

Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *Nber Macroeconomics Annual*, 4:185–216.

Campbell, J. Y. and Viceira, L. M. (1999). Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics*, 114(2):433–495.

Carrasco, M. and Florens, J.-P. (2000). Generalization of gmm to a continuum of moment conditions. *Econometric Theory*, 16(6):797–834.

Carrasco, M. and Tchuente, G. (2015). Regularized liml for many instruments. *Journal of Econometrics*, 186(2):427–442.

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

Colacito, R. and Croce, M. M. (2011). Risks for the long run and the real exchange rate. *Journal of Political Economy*, 119(1):153–181.

Davis, R., Matsui, M., Mikosch, T. V., and Wan, P. (2018). Applications of distance correlation to time series. *Bernoulli*, 24:3087–3116.

Domínguez, M. A. and Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, 72(5):1601–1615.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments. *Econometrica*, 69(5):1161–1191.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Epstein, L. G. and Zin, S. E. (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99(2):263–286.

Escanciano, J. C. (2018). A simple and robust estimator for linear regression models with strictly exogenous instruments. *Econometrics Journal*, 21(1):36–54.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica*, 45(4):939–953.

Hahn, J. (2002). Optimal inference with many instruments. *Econometric Theory*, 18(1):140–168.

Hahn, J., Hausman, J., and Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher-order bias and mse approximations. *Econometrics Journal*, 7(1):272–306.

Hall, R. E. (1988). Intertemporal substitution in consumption. *Journal of Political Economy*, 96(2):339–357.

Hansen, C., Hausman, J., and Newey, W. K. (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, 26(4):398–422.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hansen, L. P. and Singleton, K. J. (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy*, 91(2):249–265.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., and Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, 3(2):211–255.

Hsu, S.-H. and Kuan, C.-M. (2011). Estimation of conditional moment restrictions without assuming parameter identifiability in the implied unconditional moments. *Journal of Econometrics*, 165(1):87–99.

King, R. G. and Rebelo, S. (1990). Public policy and economic growth: Developing neoclassical implications. *Journal of Political Economy*, 98(5):126–150.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803.

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica*, 73(4):1103–1123.

Kunitomo, N. (2012). An optimal modification of the liml estimation for many instruments and persistent heteroscedasticity. *Annals of the Institute of Statistical Mathematics*, 64(5):881–910.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048.

Morimune, K. (1983). Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica*, 51(3):821–841.

Nelson, C. R. and Startz, R. (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica*, 58(4):967–976.

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

Phillips, G. D. A. and Hale, C. (1977). The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review*, 18(1):219–228.

Politis, D. N. and Romano, J. P. (1994). Limit theorems for weakly dependent hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*, 4(2):461–476.

Prudnikov, A. P., Brychkov, Y. A., and Marichev, O. I. (1986). *Integrals and series.* Gordon and Breach, New York.

Resnick, S. (1999). *A Probability Path.* Birkhäuser, Boston, MA.

Shao, X. and Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association*, 109(507):1302–1318.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1236–1265.

Székely, G. J. and Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.*, 42(6):2382–2412.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794.

Vissing-Jorgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853.

Wang, X. (2021). Generalized spectral tests for high dimensional multivariate martingale difference hypotheses. *Working paper.*

Yao, S., Zhang, X., and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of The Royal Statistical Society Series B-statistical Methodology*, 80(3):455–480.

Yogo, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *The Review of Economics and Statistics*, 86(3):797–810.

Zhang, X., Yao, S., and Shao, X. (2018). Conditional mean and quantile dependence testing in high dimension. *Annals of Statistics*, 46(1):219–246.