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

Carlos Velasco*

Xuexin Wang†

August 22, 2024

Abstract

This study proposes new instrumental variable (IV) estimators for linear models utilizing a continuum of instruments. The effectiveness of the new estimation method is attributed to the unique weighting function employed in the minimum distance objective functions. The proposed estimators enjoy analytical formulas and are nuisance-parameter-free, avoiding the choice of an arbitrary number of moments or bandwidth in previous literature. They are robust to weak instruments and heteroskedasticity of unknown form. Moreover, they are robust to the high dimensionality of included and excluded exogenous variables. Further, inference drawn from these estimators is also straightforward. 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 then applied to gauge the elasticity of intertemporal substitution (EIS) in consumption, a parameter of central importance in both macroeconomics

*Departamento de Economía, Universidad Carlos III de Madrid, carlos.velasco@uc3m.es . Financial support from Ministerio de Economía y Competitividad (Spain), grant PID2020-114664GB-I00 is gratefully acknowledged.

†The Wang Yanan Institute for Studies in Economics(WISE), Xiamen University, Xiamen, Fujian, China. xuexinwang@xmu.edu.cn

and finance. For quarterly data of the U.S. 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, distinct sets of IVs, various data structures, and different data ranges.

JEL Classification: C26, C36, E44

Keywords: Endogeneity; Non-integrable weighting function; Weak identification; High-dimensional robustness; EIS in consumption

1 Introduction

A substantial body of the econometric literature has been dedicated to instrumental variable (IV) methods designed to address the issue of endogeneity within linear models. However, the effectiveness of these methods is compromised by the weak instruments problem, which occurs when the strength of instrumental relevance is inadequate. The widely used two-stage least squares (2SLS) method is particularly susceptible to this issue. According to Staiger and Stock (1997), the 2SLS and limited information maximum likelihood (LIML) estimators are inconsistent and converge instead to non-standard distributions in a $n^{-1/2}$ local-to-zero parametrization of 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 employing numerous weak instruments can enhance the estimation accuracy of the LIML and bias-corrected two-stage least square (B2SLS) estimators. However, as pointed out by Bekker and van der Ploeg (2005) and Hausman et al. (2012), the consistency of the LIML and Fuller (1977) (FULL) estimators could falter in the presence of heteroskedasticity of unknown form and weak instruments. To address this issue, Hausman et al. (2012) proposed a heteroskedasticity-robust version of the FULL (HFUL) estimator, which is based on a jackknife version of the LIML estimator, referred to as HLIM. They demonstrated that 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, in practical applications, determining an appropriate number of instruments for the standard many weak IV estimators poses an exceedingly formidable challenge. In particular, within the context of linear reduced-form setups, the asymptotic properties of these estimators critically hinge upon the intricate interplay between the number of weak instruments and the so-called concentration parameter, a measure of IV strength typically unknown in real-world scenarios. Regrettably, the reduced form specification itself often remains elusive in most instances. Consequently, these estimators necessitate that the linear combination of an increasing number of instruments approximates the reduced form sufficiently well as the sample size approaches infinity. As such, established guidelines for determining the optimal number of weak instruments are conspicuously absent in the literature. The Monte Carlo simulation results presented in this work demonstrate the discernible sensitivity in the finite-sample properties of HFUL to variations in the number of instruments.

We introduce two IV estimators that effectively utilize a full continuum of instruments, offering the distinct advantage of being free from user-chosen parameters. Notably, the proposed estimators maintain analytical formulas and possess a natural jackknife structure, resembling HLIM and HFUL. We designate the HLIM-like estimator as WCIV, as its objective function involves a weighted continuum of IVs. The Fuller-like variant of WCIV is labeled as WCIVF. We establish the consistency and asymptotic normality of WCIV and WCIVF in the presence of weak instruments and heteroskedasticity of unknown form. Inference derived from these estimators is also straightforward. Extensive Monte Carlo simulations substantiate that WCIV and WCIVF consistently outperform HFUL and other competitive estimators across a wide range of scenarios. Subsequently, we employ WCIV and WCIVF to estimate the EIS in consumption, using macroeconomic datasets from the U.S. For the quarterly data ranging from Q4 1955 to Q1 2018, the WCIV and WCIVF estimates of the EIS are significantly above one and statistically different from zero. These findings hold over model transformations, distinct sets of IVs, various data structures, and different data ranges.

This study contributes in two primary ways. Firstly, it provides an enhanced methodology for estimating linear models characterized by the presence of (many) weak instruments and heteroskedasticity of unknown form. In this context, the challenge of selecting an appropriate number of IVs for the standard many weak IV estimators proves to be exceptionally daunting. The uniqueness of our approach lies in its utilization of a novel non-integrable weighting function in the minimum distance objective functions. This weighting function enjoys several attractive features which have important implications for estimation efficiency and robustness. One of the outstanding features is that its moment weights within a neighborhood of the origin tend to infinity. This is extremely important in terms of estimation efficiency, as the sample moments generated from the continuum of IVs are most informative within this vicinity. We theoretically demonstrate that WCIV is only slightly less efficient than 2SLS under some classical regularity conditions. Moreover, this weighting function is an increasing function of the dimension of included and excluded exogenous variables, indicating a "bless of dimensionality". We rigorously show that our approach can always identify the parameters as the dimension of exogenous variables goes to infinity. This feature is also important because, in the presence of weak instruments, it is advantageous to incorporate more excluded exogenous variables to augment the IV relevance, and incorporate 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 estimators enjoy analytical forms; therefore, they are easily computable. Lastly, under this weighting function, the objective functions enjoy a 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 WCIV and WCIVF estimates of the EIS in consumption offer a promising resolution to a persistent discrepancy observed between the EIS values in various model calibrations and the estimates from macroeconomic datasets. Many theoretical model calibrations necessitate significantly large EIS values to produce results that align with the stylized facts of macroeconomic dynamics. However, the majority of the previous empirical studies, such as Hall (1988), Campbell (2003), Yogo (2004) and Ascari et al. (2021), consistently report relatively small EIS values. We argue that the linear reduced forms assumed by these studies may be debatable. Existing empirical evidence suggests that linear serial dependence is not significantly present in asset returns and consumption growth at macro level. Therefore, these EIS estimates may be severely biased.

A continuum of instruments (moments) has been utilized in consistent specification tests for models defined by conditional moment restrictions; see Bierens (1982), Bierens (1990) and Bierens and Ploberger (1997), among others. Likewise, a continuum of moments has been utilized in estimation procedures for models defined by conditional moment restrictions, such as

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 akin to the one adopted in this study. However their minimum distance objective functions employ integrable weighing functions, and their IV estimators generally exhibit inferior or comparable performance to HFUL, as observed in Antoine and Lavergne (2014), and fall short of the performance achieved by WCIV and WCIVF, as demonstrated in this study. In particular, we illustrate that their approach fails to identify parameters when the dimension of exogenous variables goes to infinity. In contrast, Carrasco and Florens (2000) have introduced an optimal estimation framework involving a continuum of moments, extending the generalized method of moments (GMM) introduced by Hansen (1982). In the pursuit of estimation efficiency, their minimum distance objective function employs a random weighting function, which is analogous to the optimal weighting matrix in GMM. This approach depends on a regularization of the optimal covariance operator to address an ill-posed estimation problem which we avoid.

The remainder of the paper is organized as follows. Section 2 introduces the model setup and the new IV estimators. We provide simple analytical formulas for WCIV and WCIVF, and a valid Wald test statistic for parametric inference. Section 3 introduces the non-integrable weighting function and the minimum distance objective functions. Section 4 establishes the asymptotic theory of our proposed IV estimators. Section 6 conducts a comprehensive Monte Carlo simulation study. Section 7 presents the application of estimating EIS in consumption. Section 8 concludes. The proofs are presented in the Appendix.

Throughout the paper, the imaginary unit is $i = \sqrt{-1}$. 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.). $\vartheta_{\min}(\mathbf{A})$ denotes the smallest eigenvalue of a symmetric matrix $\mathbf{A}$. Throughout, let C denote a generic positive constant that may be different in different uses.

2 Model Setup, A Continuum of IVs and New IV Estimators

The model we focus on is defined as follows:

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

where $\mathbf{Y}_t$ is a $p \times 1$ vector of regressors. The model under the true parameter is

$$y_t = \alpha_0 + \boldsymbol{\beta}_0'\mathbf{Y}_t + \varepsilon_{0t},\ t = 1, ..., n.$$

Potentially, $\mathbf{Y}_t$ is correlated with ε_{0t}. The IV approach posits the existence of a $q \times 1$ dimensional vector of exogenous variables $\mathbf{X}_t$ (excluding a constant), $q \geq p$, such that, almost surely (a.s.)

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

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. The instrumental exogeneity condition (1) is a conditional moment restriction that is frequently encountered in macroeconomic and financial econometric models. Examples of such models include log-linearized Euler equations in asset pricing models, dynamic panel data models, and new Keynesian Phillips curves, among others.

With the exogeneity condition being satisfied, the formal identification of the parameter $\boldsymbol{\beta}_0$ hinges upon the conditional expectation $E\left(\mathbf{Y}_t|\mathbf{X}_t\right)$. In this context, consider two distinct parameter $(\alpha_1, \boldsymbol{\beta}_1')'$ and $(\alpha_2, \boldsymbol{\beta}_2')'$ values; 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

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

It is evident that the identification strength of $\boldsymbol{\beta}_0$ is directly contingent on the nature of $E\left(\mathbf{Y}_t|\mathbf{X}_t\right)$, while α_0 is always strongly identified. When $E\left(\mathbf{Y}_t|\mathbf{X}_t\right)$ flattens to zero as the sample size increases, as indicated by Assumption 1 in Section 4, the IV estimate of $\boldsymbol{\beta}_0$ may suffer from the weak identification problem.

This study employs a continuum of instruments defined as:

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

and the continuum of unconditional moment restrictions based on it has the form

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

where $\mu_y = E\left(y_t\right)$ and $\boldsymbol{\mu}_Y = E\left(\mathbf{Y}_t\right)$. It is noteworthy that the parameter α_0 is eliminated. Clearly, there exists an equivalence between (1) and (2) following Stinchcombe and White (1998).

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

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

and

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

Let $\mathbf{D}$ be a square matrix of size n, with D_{jk} representing the (j,k)th element, defined as

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

The WCIV estimator is given as

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

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

where $\mathbf{I}_n$ is an identity matrix of size n, and $\hat{\lambda}_{WCIV}$ 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 aligns with the WCIV estimator (3), but replaces $\hat{\lambda}_{WCIV}$ by

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

Clearly, WCIV and WCIVF resemble HLIM and HFUL, respectively, as $D_{jj} = 0$ for $j = 1, ..., n$. It is worth recalling that conventional k-class IV estimators take the form

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

where $\mathbf{Y}^* = [\boldsymbol{\iota}, \mathbf{Y}]$, with $\boldsymbol{\iota}$ representing a vector of ones, and $\mathbf{P}$ is a matrix that depends on an $n \times m$ matrix $\mathbf{Z}$ of instrumental variable observations with $rank(\mathbf{Z}) = m \geq p+1$. 2SLS corresponds to $\mathbf{P} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, with $\hat{\lambda} = 0$; JIVE corresponds to $\mathbf{P} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' - diag\left(\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right)$ and $\hat{\lambda} = 0$; LIML corresponds to $\mathbf{P} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and $\hat{\lambda}$ that 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 $\mathbf{P} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' - diag\left(\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right)$ and $\hat{\lambda}$ that 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 continuously differentiable 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

$$\begin{aligned}
\hat{\boldsymbol{\Omega}}_1(\boldsymbol{\theta}, \lambda) &= \frac{1}{n^3}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\varepsilon_l(\boldsymbol{\theta})^2\tilde{D}_{jl}(\lambda)\tilde{D}_{kl}(\lambda) \\
\hat{\boldsymbol{\Omega}}_2(\boldsymbol{\theta}, \lambda) &= \frac{1}{n^5}\sum_{l=1}^{n}\varepsilon_l(\boldsymbol{\theta})^2\left(\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_j\tilde{D}_{jk}(\lambda)\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_j'\tilde{D}_{jk}(\lambda)\right) \\
\hat{\boldsymbol{\Omega}}_3(\boldsymbol{\theta}, \lambda) &= \frac{1}{n^4}\left(\sum_{j=1}^{n}\sum_{k=1}^{n}\varepsilon_k(\boldsymbol{\theta})^2\tilde{\mathbf{Y}}_j\tilde{D}_{jk}(\lambda)\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_j'\tilde{D}_{jk}(\lambda)\right),
\end{aligned}$$

further

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

and

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

For $\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right) = \left(\hat{\boldsymbol{\theta}}_{WCIV},\hat{\lambda}_{WCIV}\right)$ or $\left(\hat{\boldsymbol{\theta}}_{WCIVF},\hat{\lambda}_{WCIVF}\right)$, the Wald test statistic is constructed as

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

where $\mathbf{G}(\boldsymbol{\beta}) = \partial\mathbf{g}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$. The inclusion of $\hat{\lambda}$ in these estimators follows a similar methodology as employed by Hausman et al. (2012) and is anticipated to yield enhanced finite-sample properties.

3 Minimum Distance Objective Functions and Non-integrable Weighting Function

In the previous section, we have introduced the WNIV and WNIVF estimators, which are formulated from the continuum of moments (2). In this section, we introduce their minimum distance objective functions based on a unique non-integrable weighting function. Denote

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

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 $\tilde{y}_t = y_t - \bar{y}$. An IV estimator can be attained by minimizing the sample analog of the following distance measure:

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

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

Undoubtedly, the function $W(\cdot)$ plays a pivotal role in the pursuit of estimation efficiency of the IV estimator, functioning in a manner akin to the weighting matrix in the GMM objective function outlined by Hansen (1982). As such, higher weighting values should be allocated to more informative sample moments in the minimum distance objective function. It can be shown, under some regularity conditions,

$$E\left(\left|\sqrt{n}h_n\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)\right|^2\right)\to E\left[\varepsilon_{0t}^2\left(1+\left|E\left(\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right)\right|^2-2\cos\left(\left\langle\boldsymbol{\tau},\mathbf{X}_t-\mathbf{X}_t^+\right\rangle\right)\right)\right].$$

Clearly, as $\|\boldsymbol{\tau}\|\to 0$, $E\left(\left|\sqrt{n}h_n\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)\right|^2\right)\to 0$. Hence, weighting values as high as possible in a neighborhood of the origin are extremely preferable. To this end, we adopt a non-integrable weighting function, such that

$$W(\boldsymbol{\tau})=\frac{(q-1)!!}{(2\pi)^{q/2}\|\boldsymbol{\tau}\|^{q+1}}, \tag{7}$$

where $q!!$ is the double factorial,

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

One outstanding feature of (7) is that its values tend to infinity as $\|\boldsymbol{\tau}\|\to 0$, being substantially different from the standard normal density function employed in Escanciano (2018) and Antoine and Lavergne (2014). Figure 1 illustrates 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 for fixed $\boldsymbol{\tau}$. Note that, by applying an approximation to $(q-1)!!$, when $q>1$, we have

$$W(\boldsymbol{\tau})\approx\frac{c\sqrt{e}}{\|\boldsymbol{\tau}\|}\left(\frac{q-1}{2\pi e\|\boldsymbol{\tau}\|^2}\right)^{q/2},$$

where $c=\sqrt{\pi}$ when $q-1$ is even, and $c=\sqrt{2}$ when $q-1$ is odd. Therefore, for a fixed value $\|\boldsymbol{\tau}\|$ in a neighborhood of the origin, values of $W(\boldsymbol{\tau})$ increase as q increases, indicating a "bless of dimensionality" in terms of estimation efficiency. In contrast, a q-dimensional standard normal density function is a decreasing function of q, for given $\|\boldsymbol{\tau}\|$. Notably, its value equals $(2\pi)^{-q/2}$

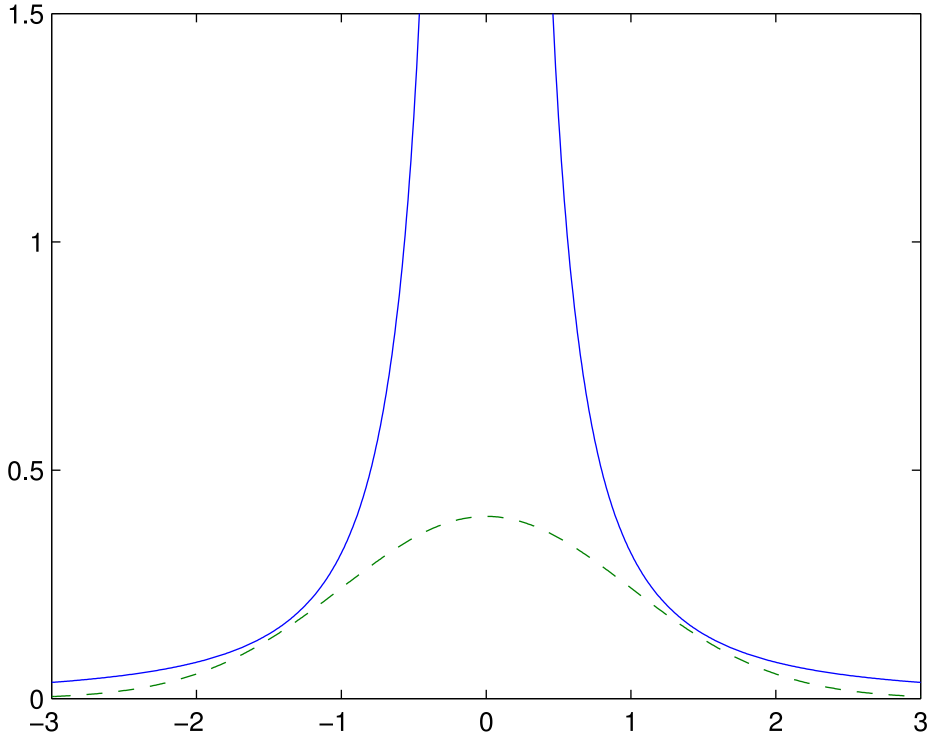

Figure 1: Standard normal density function (dashed curve) vs. $1/(\pi\boldsymbol{\tau}^2)$ (solid curve)

at the origin, being the maximum, and shrinks to zero when q increases.

The non-integrable 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 (2024) in a testing framework. In the subsequent discussion, we write

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

where $\boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) = \frac{(q-1)!!}{(2\pi)^{q/2}\|\boldsymbol{\tau}\|^{q+1}} d\boldsymbol{\tau}$ for notational simplicity.

The third feature associated with the non-integrable weighting function is that $\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta},\boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right)$ enjoys a convenient analytical form, as demonstrated by Lemma 3.1.

Lemma 3.1 *When* $E\left(\varepsilon_t^2\right) < \infty, E\left\|\mathbf{X}_t\right\|^2 < \infty$,

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

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

Proof. See the Appendix. ■

Beneath the specialties of the non-integrable weighting function (7) lies a fundamental property regarding parameter identification, as demonstrated in the following proposition:

Proposition 3.1 *When* $E\left(\varepsilon_{0t}^2\right) < \infty, E\left\|\mathbf{X}_t\right\|^2 < \infty$, *for* $q \to \infty$,

$$E\left(\varepsilon_{0t}|\mathbf{X}_t\right) = 0 \ \ a.s.$$

if and only if

$$E\left[\varepsilon_{0t}\varepsilon_{0t}^+ \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right] = 0 \ \ and \ E\left(\varepsilon_{0t}\right) = 0.$$

where $\left(\varepsilon_{0t}^+, \left(\mathbf{X}_t^+\right)'\right)'$ *is an i.i.d. copy of* $(\varepsilon_{0t}, \mathbf{X}_t')'$.

Proof. See the Appendix. ■

This proposition indicates that $\boldsymbol{\beta}_0$ is always identifiable by $\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta}_0, \boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) = 0$, even as $q \to \infty$, implying the high-dimensional robustness of WCIV and WCIVF.

It is also worthwhile mentioning that it is conceivable to introduce a random weighting function following Carrasco and Florens (2000). However, this approach becomes extremely challenging under weak instruments, heteroskedasticity of unknown form and high-dimensionality of exogenous variables, as it requires a regularization of the weighting function which introduces a nuisance parameter and proves intricate to implement in practical applications. In contrast, our approach employs a nonrandom weighting function in the distance measure, effectively circumventing the challenging issues of determining an appropriate number of instruments or a nuisance parameter.

While it is feasible to derive a new IV estimator by minimizing the sample analog of (8), Monte Carlo simulations have revealed that this estimator can be significantly biased in the presence of weak instruments and high-dimensionality of $\mathbf{X}_t$. To address this issue and improve estimation accuracy under such circumstances, the WCIV objective function is constructed like a LIML objective function such that:

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

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

Subsequently, the WCIV estimator is defined as:

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

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

The derivation of the WCIV estimator formula (3) is straightforward and analogous to the derivation for HLIM. Moreover, WCIV remains invariant to normalization, similar to HLIM. Further, in order to guard against the potential moment problem of WCIV, we propose the WCIVF estimator, following the approach presented in Fuller (1977), Hahn et al. (2004), and Hausman et al. (2012).

3.1 Comparison with Antoine and Lavergne (2014)

It is worth mentioning that the continuum of moments $E\left[(y_t - \boldsymbol{\theta}_0' \mathbf{Y}_t^*) \exp(i \langle \boldsymbol{\tau}, \mathbf{X}_t \rangle)\right] = 0$, for all $\boldsymbol{\tau} \in \mathbb{R}^q$, is also employed in Antoine and Lavergne (2014). Using a standard normal density function in the objective function, their minimum distance (MD) estimator is calculated as

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

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

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

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

Correspondingly, we have

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

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

where $\hat{\lambda}_{WMD}$ is the minimum value of the objective function, which can be explicitly computed as the smallest eigenvalue of $\left(\check{\mathbf{Y}}^{*\prime}\check{\mathbf{Y}}^{*}\right)^{-1}\check{\mathbf{Y}}^{*\prime}\mathbf{K}\check{\mathbf{Y}}^{*}$ with $\check{\mathbf{Y}}^{*}=[\mathbf{y},\mathbf{Y}^{*}]$. The Fuller-style variant of WMD (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 may appear that both WMD and WCIV (WMDF and WCIVF) share many similarities, but they are constructed on distinct estimation frameworks. In particular, WMD and WMDF are not robust to the high-dimensionality of $\mathbf{X}_t$. To see this point, note that the minimum distance objective functions of WMD and WMDF are based on the population moment

$$E\left[\left(y_t-\boldsymbol{\theta}'\mathbf{Y}_t^{*}\right)\left(y_t^{+}-\boldsymbol{\theta}'\mathbf{Y}_t^{*+}\right)\exp\left(-\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|^2/2\right)\right].$$

By Cauchy-Schwarz inequality, it is easy to obtain

$$\begin{aligned}
&\left|E\left[\left(y_t-\boldsymbol{\theta}'\mathbf{Y}_t^{*}\right)\left(y_t^{+}-\boldsymbol{\theta}'\mathbf{Y}_t^{*+}\right)\exp\left(-\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|^2/2\right)\right]\right|\\
\leq\ & E\left(\left(y_t-\boldsymbol{\theta}'\mathbf{Y}_t^{*}\right)^2\right)\left[E\left(\exp\left(-\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|^2\right)\right)\right]^{1/2}\\
\leq\ & C\left(\boldsymbol{\theta}\right)\left[E\left(\exp\left(-\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|^2\right)\right)\right]^{1/2},
\end{aligned}$$

because under some regularity conditions, $E\left(\left(y_t-\boldsymbol{\theta}'\mathbf{Y}_t^{*}\right)^2\right)$ is finite for each $\boldsymbol{\theta}$. As $q\to\infty$, $E\left(\exp\left(-\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|^2\right)\right)\to 0$. To further understand this, assuming the components of $\mathbf{X}_t=(X_{1t},...,X_{qt})$ are i.i.d., then

$$E\left(\exp\left(-\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|^2\right)\right)=\left[E\left(\exp\left(-\left(\mathbf{X}_{1t}-\mathbf{X}_{1t}^{+}\right)^2\right)\right)\right]^{q}.$$

Since $0<E\left(\exp\left(-\left(\mathbf{X}_{1t}-\mathbf{X}_{1t}^{+}\right)^2\right)\right)<1$, as $q\to\infty$, $E\left(\exp\left(-\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|^2\right)\right)\to 0$ exponentially fast. This result holds for any $\boldsymbol{\theta}$, implying $\boldsymbol{\theta}_0$ will not be identified when $q\to\infty$. In other words, strikingly different from WCIV and WCIVF, WMD and WMDF are not robust to the high-dimensionality of $\mathbf{X}_t$. The Monte Carlo simulation results reported in the Appendix illustrate that the finite-sample properties of WMD and WMDF deteriorate severely in the case that the elements of $\mathbf{X}_t$ are i.i.d. even when q is moderate, while WCIV and WCIVF maintain execllent finite-sample properties in all the cases.

4 Asymptotic Theory

To derive the asymptotic theory of new estimators, we introduce the following assumptions.

Assumption 1

$$\mathbf{Y}_t = \frac{\mathbf{R}_n \mathbf{f}(\mathbf{X}_t)}{\sqrt{n}} + \boldsymbol{\eta}_t,$$

where $\mathbf{R}_n = \tilde{\mathbf{R}} diag(r_{1,n}, ..., r_{q,n})$ *such that* $\tilde{\mathbf{R}}$ *is a* $q \times q$ *nonsingular matrix, for each* j, $r_{j,n} = \sqrt{n}$ *or* $r_{j,n}/\sqrt{n} \to 0$, $r_n = \min_{1 \le j \le q} r_{j,n} \to \infty$, *and* $\mathbf{f} : \mathbb{R}^q \to \mathbb{R}^p$ *is a measurable function.* $\frac{1}{n^2}\sum_{j,k} \tilde{\mathbf{f}}(\mathbf{X}_j) D_{jk} \tilde{\mathbf{f}}(\mathbf{X}_k)'$ *is finite and positive definite, where* $\tilde{\mathbf{f}}(\mathbf{X}_t) = \mathbf{f}(\mathbf{X}_t) - \frac{1}{n}\sum_{j=1}^n \mathbf{f}(\mathbf{X}_j)$, *for* n *sufficiently large.* $\vartheta_{\min}\left(\sum_{t=1}^n \mathbf{f}(\mathbf{X}_t)\mathbf{f}(\mathbf{X}_t)'/n\right) \ge 1/C$ *for* n *sufficiently large.*

Assumption 2 $\left\{(\varepsilon_{0t}, \mathbf{X}_t', \boldsymbol{\eta}_t')'\right\}$ *is i.i.d., such that* $E(\varepsilon_{0t}|\mathbf{X}_t) = 0, E(\boldsymbol{\eta}_t|\mathbf{X}_t) = 0$. *Further,* $E\left(\|\mathbf{X}_t\|^2\right) < C$, $\sup_t E(\varepsilon_{0t}^2|\mathbf{X}_t) < C$, $\sup_t E\left(\|\boldsymbol{\eta}_t\|^2 |\mathbf{X}_t\right) < C$, $Var\left((\varepsilon_{0t}, \boldsymbol{\eta}_t')' |\mathbf{X}_t\right) = diag(\Phi_t^*, 0)$, *and* $\vartheta_{\min}\left(\sum_{t=1}^n \Phi_t^*/n\right) \ge 1/C$ *for* n *sufficiently large.*

Assumption 1 is comparable to Assumption 2 in Hausman et al. (2012), allowing linear combinations of $\boldsymbol{\beta}$ to have different degrees of identification. It is noted that the rates of decay of the reduced form to zero are slower than $1/\sqrt{n}$, which has been labeled as nearly-weak identification or semi-strong identification by previous studies. Here we adopt the "(many) weak instruments" terminology, following Hansen et al. (2008), Newey and Windmeijer (2009), and Hausman et al. (2012). Our framework accommodates IV regressions involving included exogenous variables, see Hausman et al. (2012) for more details. Moreover, it is worthwhile mentioning that $\mathbf{f}(\mathbf{X}_t)$ is an unknown function that we do not need to estimate.

Assumption 2 imposes some independence conditions and may be extended to weakly dependent time series processes. Further, it requires bounded second conditional moments of disturbances and uniform nonsingularity of the variance of the reduced form disturbances, which are comparable to those under Assumption 3 in Hausman et al. (2012).

Theorem 4.1 establishes the consistency for WCIV and WCIVF.

Theorem 4.1 *Under Assumptions 1-2, for* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ *or* $\hat{\boldsymbol{\beta}}_{WCIVF}$, $\hat{\alpha} = \hat{\alpha}_{WCIV}$ *or* $\hat{\alpha}_{WCIVF}$,

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

Proof. See the Appendix. ■

To discuss the asymptotic normality, an additional assumption is required. Define

$$\mathbf{V}(\boldsymbol{\theta}) = \mathbf{V}_1(\boldsymbol{\theta}) + \mathbf{V}_2(\boldsymbol{\theta}) - \mathbf{V}_3(\boldsymbol{\theta}) - \mathbf{V}_3(\boldsymbol{\theta})',$$

where

$$\begin{aligned}
\mathbf{V}_1(\boldsymbol{\theta}) &= E\left((y_t - \alpha - \boldsymbol{\beta}'\mathbf{Y}_t)^2 \left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\left(\mathbf{f}\left(\mathbf{X}_t^{++}\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\| \left\|\mathbf{X}_t - \mathbf{X}_t^{++}\right\|\right) \\
\mathbf{V}_2(\boldsymbol{\theta}) &= E\left[(y_t - \alpha - \boldsymbol{\beta}'\mathbf{Y}_t)^2\right] E\left((\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}}) \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right) E\left((\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right) \\
\mathbf{V}_3(\boldsymbol{\theta}) &= E\left((y_t - \alpha - \boldsymbol{\beta}'\mathbf{Y}_t)^2 \left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right) \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right) E\left((\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right)
\end{aligned}$$

with $\boldsymbol{\mu}_{\mathbf{f}} = E\mathbf{f}(\mathbf{X}_t)$.

Assumption 3 $\sup_t E\left(\varepsilon_{0t}^4 | \mathbf{X}_t\right) < C$, $\sup_t E\left(\|\boldsymbol{\eta}_t\|^4 | \mathbf{X}_t\right) < C$ *a.s.*, $E\|\mathbf{X}_t\|^4 < \infty$ *and* $E\|\mathbf{f}(\mathbf{X}_t)\|^4 < \infty$. $\mathbf{V}(\boldsymbol{\theta}_0)$ *is positive definite.*

We state the asymptotic normality theorem.

Theorem 4.2 *Under Assumptions 1-3, for* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ *or* $\hat{\boldsymbol{\beta}}_{WCIVF}$,

$$\mathbf{R}_n' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} \mathbf{N}\left(0, \boldsymbol{\Pi}^{-1}\mathbf{V}(\boldsymbol{\theta}_0)\boldsymbol{\Pi}^{-1}\right),$$

where

$$\boldsymbol{\Pi} = -E\left((\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right).$$

Proof. See the Appendix. ■

This theorem establishes that, when suitably normalized, the asymptotic distributions of $\hat{\boldsymbol{\beta}}_{WCIV}$ and $\hat{\boldsymbol{\beta}}_{WCIVF}$ converge to the standard normal distribution. It is noteworthy to emphasize that $\mathbf{V}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Pi}$ enjoy analytical representations, but involve $\mathbf{f}(\mathbf{X}_t)$, which is unobservable.

We have introduced $\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)$ and $\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)$ in the definition of the Wald statistic. In the subsequent theorem, we establish the consistency of $\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)$ and $\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)$ to $\mathbf{V}\left(\boldsymbol{\theta}_0\right)$ and $\boldsymbol{\Pi}$ under a proper normalization, and affirm the validity of the Wald test statistic for conducting parameter inference concerning $\boldsymbol{\beta}_0$.

Theorem 4.3 *Under Assumptions 1-3, if* $\mathbf{g}\left(\cdot\right)$ *is continuously differentiable and* $\mathbf{G}\left(\boldsymbol{\beta}_0\right)$ *is of full rank, under the null (5), considering* $\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)=\left(\hat{\boldsymbol{\theta}}_{WCIV},\hat{\lambda}_{WCIV}\right)$ *or* $\left(\hat{\boldsymbol{\theta}}_{WCIVF},\hat{\lambda}_{WCIVF}\right)$,

$$n\mathbf{R}_n^{-1}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\mathbf{R}_n^{-1\prime}\xrightarrow{p}\mathbf{V}\left(\boldsymbol{\theta}_0\right),$$

$$n\mathbf{R}_n^{-1}\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1\prime}\xrightarrow{p}\boldsymbol{\Pi},$$

and

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

Proof. See the Appendix. ■

This theorem demonstrates that under the null, $W_n\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)$ has a convenient chi-squared asymptotic distribution, despite the fact that the degree of identification remains unknown. An important implication of the Wald test statistic property is that it enables large-sample inference in the usual manner, even in the absence of knowledge regarding the degree of weak identification. In particular, we can compute t-statistics by treating $\hat{\beta}_j$ as if it were normally distributed, with a mean β_{0j} and a variance $\left(\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\right)_{jj}/n$, so that under the null, the t-statistic $\left(\hat{\beta}_j-\beta_{0j}\right)/\sqrt{\left(\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\right)_{jj}/n}$ is asymptotically normal distributed. Our Monte Carlo simulations further demonstrate that these t-statistics exhibit excellent finite-sample properties across a wide range of scenarios, and in our application, we report $\sqrt{\left(\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\right)_{jj}/n}$ as if they were conventional standard errors.

5 Discussion on Efficiency of WCIV and WCIVF

Estimation efficiency is a highly desirable property of an estimator. In the conventional asymptotic framework, the efficiency of IV estimators can be achieved by employing an increasing number of instruments. While some literature, such as Hahn (2002), Anderson et al. (2010), and Kunitomo (2012), discusses estimation efficiency in the context of many instruments, these theoretical results are limited in scope. In the many weak instruments asymptotics, an increasing number of instruments is required to ensure estimation consistency. However, 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 an IV estimator involves an optimal weighting matrix, which is difficult to estimate accurately under many instruments and heteroskedasticity of unknown form.

Regarding estimation methods using a continuum of moments, intuition suggests that WCIV (WCIVF) is more efficient than WMD (WMDF) due to the specialty of the non-integrable weighting function matching the most relevant moments automatically. It is admitted that the theoretical validation is quite challenging in the case of weak IV and conditional heteroskedasticity, so we focus on a limited but still relevant framework.

We set $p \leq q \leq 2$, $E\left(\varepsilon_{0t}^2|\mathbf{X}_t\right) = \sigma_\varepsilon^2$ and first assume that $\mathbf{Y}_t$ and $\mathbf{X}_t$ are i.i.d. normally distributed with covariance $\Sigma_{\mathbf{YX}}$ and $Var\left(\mathbf{X}_t\right) = \Sigma_{\mathbf{XX}} = \sigma_{\mathbf{X}}^2\mathbf{I}_q$. Define

$$\boldsymbol{\Omega}\left(\boldsymbol{\theta}\right) = \boldsymbol{\Omega}_1\left(\boldsymbol{\theta}\right) + \boldsymbol{\Omega}_2\left(\boldsymbol{\theta}\right) - \boldsymbol{\Omega}_3\left(\boldsymbol{\theta}\right) - \boldsymbol{\Omega}_3\left(\boldsymbol{\theta}\right)',$$

in which

$$\begin{aligned}
\boldsymbol{\Omega}_1\left(\boldsymbol{\theta}\right) &= E\left(\left(y_t - \alpha - \boldsymbol{\beta}'\mathbf{Y}_t\right)^2\left(\mathbf{Y}_t^{+} - \boldsymbol{\mu}_Y\right)\left(\mathbf{Y}_t^{++} - \boldsymbol{\mu}_Y\right)'\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\left\|\mathbf{X}_t - \mathbf{X}_t^{++}\right\|\right)\\
\boldsymbol{\Omega}_2\left(\boldsymbol{\theta}\right) &= E\left[\left(y_t - \alpha - \boldsymbol{\beta}'\mathbf{Y}_t\right)^2\right]E\left(\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right)E\left(\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)'\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right)\\
\boldsymbol{\Omega}_3\left(\boldsymbol{\theta}\right) &= E\left(\left(y_t - \alpha - \boldsymbol{\beta}'\mathbf{Y}_t\right)^2\left(\mathbf{Y}_t^{+} - \boldsymbol{\mu}_Y\right)\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right)E\left(\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)'\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right),
\end{aligned}$$

and

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

We demonstrate in the Appendix that under these regularity conditions, the asymptotic variance of WCIV is

$$\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{\theta_0}\right)\boldsymbol{\Upsilon}^{-1} = ARE_{WCIV}\left(q\right)\cdot\sigma_\varepsilon^2\left(\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\right)^{-1},$$

where $ARE_{WCIV}\left(q\right)$, the asymptotic relative efficiency of WCIV with respect to 2SLS, is equal to $ARE_{WCIV}(1) = \frac{\pi}{3} \approx 1.047$ and $ARE_{WCIV}(2) \approx 1.035$, independently of any other parameter beyond q, and in particular of $\sigma^2_{\mathbf{X}}$. This indicates that WCIV is only slightly less efficient than 2SLS in this set up and that the efficiency loss shrinks with q.

In contrast, by evaluating appropriate versions of $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Omega}$ for a Gaussian kernel, the asymptotic relative efficiency of WMD with respect to 2SLS depends on q and $\sigma^2_{\mathbf{X}}$, and satisfies

$$\begin{aligned} ARE_{WMD}\left(1, \sigma^2_{\mathbf{X}}\right) &= \frac{\left(1 + 2\sigma^2_{\mathbf{X}}\right)^3}{\left(1 + \sigma^2_{\mathbf{X}}\right)^{3/2}\left(1 + 3\sigma^2_{\mathbf{X}}\right)^{3/2}} \in (1,\ 1.540) \\ ARE_{WMD}\left(2, \sigma^2_{\mathbf{X}}\right) &\in (1,\ 1.778)\,, \end{aligned}$$

for $q = 1, 2$, where inefficiency increases with q and the lower bound is only approached as $\sigma^2_{\mathbf{X}} \to 0$. However, for usual values of $\sigma^2_{\mathbf{X}}$, WMD can be severely inefficient compared to 2SLS or WCIV.

Further, we also show that WMD is uniformly less efficient than WCIV in the case of exponentially distributed data for $q = 1$, though now the differences can be smaller than in the Gaussian case. These findings echo the specialties of the non-integrable weighting function employed in the objective function of WCIV.

6 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. Following Hausman et al. (2012), the instruments used in HFUL are

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

where $\mathbf{X}_t^r = \left(X_{1,t}^r, ..., X_{q,t}^r\right)'$, in which r is a positive integer, $d_l \in \{0, 1\}$, and $\Pr\left(d_l = 1\right) = 1/2$. We consider $L = 1$, 4, or 9, that is, when $L = 1$, the instruments are $(1, \mathbf{X}_t')'$; when $L = 4$, $(1, \mathbf{X}_t', (\mathbf{X}_t^2)', (\mathbf{X}_t^3)', (\mathbf{X}_t^4)')'$; 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 among these estimators are in terms of median biases, ranges between the 0.05 and 0.95 quantiles, and empirical rejection frequencies for t-statistics at the 5% nominal level. The number of Monte Carlo simulations is $10,000$.

6.1 Setup 1

The models that we consider are

$$M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t},\ Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t,$$

$$M_2 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t},\ Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t}^2 + \eta_t,$$

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

where $1\{\cdot\}$ denotes an indicator function. In M_1, $E(Y_t|\mathbf{X}_t)$ is linear. In M_2 and M_3, $E(Y_t|\mathbf{X}_t)$ is nonlinear. To mimic empirical situations, the elements of $\mathbf{X}_t = (X_{1,t}, ..., X_{q,t})'$ follow

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

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

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

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 $\phi \neq 0$. We set $\phi = 0,\ 0.5$. 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 $q = 3, 10, 15$ to mimic the application of the EIS estimation.

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 produce accurate empirical sizes, while HFUL1 is undersized and HFUL9 is oversized, especially for the 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 the high-dimensional cases.

2. For M_2, HFUL1 is severely median biased and dispersed, while HFUL4 and HFUL9 are 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 the empirical rejection frequencies for the t-statistics well controlled. In terms of DecR, WCIV and WCIVF outperform WMD and WMDF substantially, and are 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 are 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

		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	0.0500	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 $M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t}$, $Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t$. 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.

rejection frequencies of the t-statistics are close to the nominal value. They are considerably less dispersed than WMD and WMDF in all cases, whose variance increases with q unlike WCIV and WCIVF for M_2 and M_3 as predicted by our theory. In comparison with HFUL, both WCIV and WCIVF exhibit less dispersion in most 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 illustrate that HFUL may provide misleading estimates when the reduced forms are not well-approximated using linear combinations of the selected instruments.

		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	0.0550	0.0780	0.5172	0.0302	0.0850

Table 2: Linear IV model $M_2 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t}$, $Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t}^2 + \eta_t$. 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.

6.2 Setup 2

The alternative linear models, which are similar to Antoine and Lavergne (2014), are

$$M_4 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_{0t},\ Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \eta_t,$$

$$M_5 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_{0t},\ Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \exp(0.5 + 0.5X_{1,t,})\eta_t,$$

$$M_6 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_{0t},\ Y_t = \exp\left(\frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t}\right) + \eta_t.$$

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$\phi = 0$					$q = 3$			
	Med	-0.0062	-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	1.1680	2.3200	1.7693	1.3057	2.0461	3.8185
	Rej	0.0459	0.0460	0.0471	0.0585	0.0075	0.0328	0.0806

Table 3: Linear IV model $M_3 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t}$, $Y_t = 1\left\{\sqrt{\frac{c/q}{n}}\sum_{j=1}^{q} X_{j,t} + \eta_t > 0\right\}$. 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.

Note that heteroskedasticity in model disturbances is allowed for. In all models ε_{0t} 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 setup 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 set $c = 4$, 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($C =$

		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	0.9230	0.7440	0.8893	1.1267
	Rej	0.0506	0.0506	0.0548	0.0570	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	0.0560	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 $M_4 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_{0t}$, $Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \eta_t$. 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.

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, outperforming other alternatives, especially when the q values are large and the reduced forms are nonlinear. 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.

7 Application to Estimating the EIS in Consumption

In this section, we use WCIV and WCIVF to estimate the EIS in consumption for macroeconomic datasets from the U.S. The EIS in consumption, which measures how much consumers change

		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	0.0620	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	0.1140	0.1557

Table 5: Linear IV model $M_5 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_{0t},\ Y_t = \frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t} + \exp(0.5 + 0.5X_{1,t,})\eta_t$. 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.

their expected consumption growth rate in response to changes in the expected return on any asset, is a parameter of central importance in macroeconomics and finance. For instance, King and Rebelo (1990) demonstrates that the EIS is the key parameter in a simple neoclassical model of endogenous growth, which involves taxation. In consumption-based asset pricing models, the EIS determines the optimal consumption rule, as observed in Campbell and Viceira (1999).

Sequentially, the 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

		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	0.0560	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	0.0490	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	0.0610	0.0640	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 $M_6 : y_t = \alpha_0 + \beta_0 Y_t + \sqrt{0.5 + 0.5X_{1,t}^2}\varepsilon_{0t}$, $Y_t = \exp\left(\frac{\sqrt{c/q}}{n^{0.45}} \sum_{j=1}^{q} X_{j,t}\right) + \eta_t$. 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.

2. However, to date, empirical results based on macroeconomic datasets 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 did not appropriately consider the time-aggregation problem of the data, and the employed instruments were problematic. When valid instruments are employed, Hall (1988) finds that the 2SLS estimates of the EIS for the U.S. 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

[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.

et al. (2021) employ weak-instrument-robust inference procedures on macroeconomic datasets, following Staiger and Stock (1997), Kleibergen (2002), Moreira (2003) and Kleibergen (2005); however, they reach similar conclusions as in Hall (1988).

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, γ be the coefficient of relative risk aversion, and $\theta = (1-\gamma)/(1-1/\psi)$, where ψ is the EIS in consumption. 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

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

and

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

with

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

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.

For macroeconomic datasets from the U.S., researchers have commonly employed instrumental variable regression techniques to gauge the EIS. Typically, these methodologies require selecting an IV set $\mathbf{X}_t$, drawn from the elements within the information set. By invoking the law of iterated expectations, the ensuing expressions

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

and its reverse form

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

can be derived. A majority of the empirical investigations, such as Hall (1988), Campbell (2003), Yogo (2004), Beeler and Campbell (2012), and Ascari et al. (2021), has posited linear reduced forms, thereby specifying the moment conditions as follows:

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

and

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

It is worth noting that previous studies have employed the lag terms of asset returns, consumption growth, and other related macroeconomic variables as instruments. In particular, Yogo (2004) incorporates the lag terms of the nominal asset return, inflation rate, consumption growth, and log dividend-price ratio, while Campbell (2003) includes the lag terms of the real asset return, real consumption growth, and log dividend-price ratio. Additionally, Beeler and Campbell (2012) utilize the lag terms of the real interest rate, real stock return, real consumption growth, and log dividend-price ratio. On the other hand, Ascari et al. (2021) consider the lag terms of the real consumption growth and real asset return.

However, the linear reduced forms assumed by these studies may be debatable. Existing empirical evidence suggests that linear serial dependence is not significantly present in asset returns and consumption growth at macro level. Further, the weak instrument evidence, as indicated by the relatively low first-stage F-statistic values reported in Yogo (2004) and Ascari et al. (2021), might be attributed to the misspecification of the reduced forms. While aforementioned empirical studies obtain small EIS results, Carrasco and Tchuente (2015) and Escanciano (2018) obtain relatively larger EIS estimates by adopting nonlinear instruments in their methods. However, despite the magnitudes of these estimates, they fail to achieve statistical significance when compared to zero. It is noteworthy that both aforementioned methods are not robust to weak instruments and heteroskedasticity of unknown form. As such, it seems promising to apply the methodology developed in this study to gauge the EIS in consumption.

7.1 The U.S. Quarterly Data in Ascari et al. (2021)

We first utilize the data set from Ascari et al. (2021) which includes the quarterly data on equity markets and macroeconomic variables from Q4 1955 to Q1 2018. For the nominal interest rate i_t, this analysis employs the three-month treasury bill rate; the nominal stock return s_t is the S&P 500 return. c_t is the log of the real consumption of nondurables and services, 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 $s_t - \pi_{t+1}$ are considered in the empirical analysis. The EIS is estimated using the real interest rate as the asset return.[3] $\mathbf{X}_t$ comprises the lag terms of the real interest rate, real stock return, consumption growth, and the first difference of the log dividend-price ratio ($\triangle dp_t$). 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.[4] 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). For comparison, the estimates obtained using alternative estimation procedures are also reported. For HFUL, the instruments include constant and pairwise instruments $\left(\mathbf{X}_t', (\mathbf{X}_t^2)', \mathbf{X}_t' d_1, ..., \mathbf{X}_t' d_{L-2}\right)'$, where $d_l \in \{0, 1\}$, $\Pr(d_l = 1) = 1/2$. We consider $L = 1$, 2 or 6, i.e, when $L = 1$, the instruments are $(1, \mathbf{X}_t')'$; when $L = 2$, $(1, \mathbf{X}_t', (\mathbf{X}_t^2)')'$; when $L = 6$, $(1, \mathbf{X}_t', (\mathbf{X}_t^2)', \mathbf{X}_t' d_1, ..., \mathbf{X}_t' d_4)'$. We denote these HFUL estimates as HFUL1, HFUL2, and HFUL6, respectively. We have to utilize a smaller number of instruments to avoid singular matrix problems in HFUL. We set $C = 1$ for WCIVF, WMDF, and HFUL.

Table 7 presents the estimation results for $1/\psi$ and ψ. Notably, the WCIV and WCIVF estimates of the EIS (ψ) appear to be large, and statistically significant at the 10% significant level. These findings hold true over model transformation, with the inverse of the WCIV and WCIVF estimates of $1/\psi$ aligning with those of ψ. The WMD and WMDF estimates of the EIS are comparable to those of WCIV and WCIVF in some cases, indicating the robustness of the exploitation of the continuum of instruments. However, the WMD estimates are not statistically significant at the 10% significant level, and the WMDF estimates of ψ are generally much smaller than the WCIV and WCIVF estimates. Furthermore, the WMD and WMDF estimates of ψ frequently differ substantially. For example, for the first IV set, the WMD estimate of the EIS is 1.19 and statistically insignificant at the 10% significant level, while the WMDF estimate is 0.77, and statistically significant at the 10% significant level. The HFUL estimates of the EIS are

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

[4] 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.

	WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL2	HFUL6	2SLS
				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: The estimates of the EIS using real interest rate as the asset return. The quarterly data range is from Q4 1955 to Q1 2018. The 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. $*$ and $**$ represent the significance at 10% and 5% respectively.

generally greater than the 2SLS estimates, but much less than the WCIV and WCIVF estimates.

7.2 The U.S. 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 the monthly CRSP NYSE/AMEX Value-weighted Indices. The real interest rates and real stock returns are ex-ante. See the appendix of Beeler and Campbell (2012) for a detailed description of the data, sources, and transformation used.

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

	WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL2	HFUL6	2SLS
				Set 1				
ψ	1.21**	1.19**	1.10*	0.83**	0.30**	0.29**	0.27**	0.29**
	(0.61)	(0.59)	(0.59)	(0.37)	(0.13)	(0.13)	(0.14)	(0.13)
$1/\psi$	0.83*	0.81*	0.91*	0.57**	2.90**	2.91**	2.97**	1.13***
	(0.46)	(0.45)	(0.50)	(0.26)	(1.32)	(1.32)	(1.46)	(0.31)
				Set 2				
ψ	0.90**	0.87**	0.95**	0.34***	0.30**	0.29**	0.27*	0.29***
	(0.36)	(0.35)	(0.48)	(0.12)	(0.13)	(0.13)	(0.15)	(0.12)
$1/\psi$	1.11*	1.02*	1.06*	0.19**	2.89**	2.90**	2.82**	1.12***
	(0.65)	(0.56)	(0.63)	(0.07)	(1.31)	(1.32)	(1.32)	(0.31)
				Set 3				
ψ	1.21*	1.19*	1.10*	0.84**	0.30**	0.30**	0.32**	0.29**
	(0.61)	(0.60)	(0.60)	(0.38)	(0.13)	(0.14)	(0.14)	(0.13)
$1/\psi$	0.83*	0.81*	0.91*	0.59**	2.89**	2.86**	2.71**	1.13***
	(0.46)	(0.45)	(0.50)	(0.27)	(1.31)	(1.30)	(1.38)	(0.31)

Table 8: The 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 the 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 the first difference of the log price-dividend ratio. The third set consists of the first lag terms of real interest rate, real stock return, consumption growth, and the first difference of the log price-dividend ratio. The values in the brackets are the standard deviations of the corresponding estimates. $*$ and $**$ represent the significance at 10% and 5% respectively.

consists of the lag terms of the real interest rate, real stock return, consumption growth, and the first difference of the 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 estimates of the EIS are also quite large, although the data range, data structure, and $\mathbf{X}_t$ are different.

In summary, we obtain large WCIV and WCIVF estimates of the EIS in consumption, which well exceed one, and are statistically significant from zero. Further, these findings are robust to the distinct sets of $\mathbf{X}_t$, model transformations, data structures and ranges. These results are strikingly different from those of HFUL and lend strong support to the practices of some model calibrations, which choose substantially large values of the EIS.

8 Conclusion

This study proposes two novel IV estimators, namely WCIV and WCIVF, utilizing a continuum of instruments and a unique non-integrable weighting function in the minimum distance objective functions 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 macroeconomic datasets of the U.S. Our results show the WCIV and WCIVF estimates well exceed one and are statistically significant, being strikingly different from the results obtained using alternative approaches.

9 Appendix

Denote $\sum_{j,k} = \sum_{j=1}^{n} \sum_{k=1}^{n}$. Let w.p.a.1 denote with probability approaching one.

Lemma 9.1 *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:* $lim_{\varepsilon \to 0} \int_{\mathbb{R}^q \setminus \{\varepsilon B + \varepsilon^{-1} B^c\}}$*, where* B *is the unit ball centered at* 0 *and* B^c *is the complement of* B*.*

Proof. See Székely and Rizzo (2005) for a proof. ■

Proof of Lemma 3.1 . By Cauchy-Schwarz inequality,

$$\int_{\mathbb{R}^q} |h(\boldsymbol{\beta}, \boldsymbol{\tau})|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) = \int_{\mathbb{R}^q} \left|E\left[(\varepsilon_t - E(\varepsilon_t))\left(\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle) - E\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right)\right]\right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau})$$
$$\leq E\left[(\varepsilon_t - E(\varepsilon_t))^2\right] E\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\| < \infty.$$

Now

$$
|h(\boldsymbol{\beta}, \boldsymbol{\tau})|^2 = E\left[(\varepsilon_t - E(\varepsilon_t)) \exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right] E\left[(\varepsilon_t - E(\varepsilon_t)) \exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right]
$$
$$
= -E\left[(\varepsilon_t - E(\varepsilon_t))\left(\varepsilon_t^+ - E(\varepsilon_t)\right)\left(1 - \exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\right].
$$

Then by Fubini's Theorem and Lemma 9.1, we obtain (8). ■

Proof of Proposition 3.1 . We prove the case that $q = \infty$ with the understanding that the inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$ are defined on $\mathbb{R}^\infty$. It is noted that the separable Hilbert space $H = (\mathbb{R}^\infty, \langle \cdot, \cdot \rangle)$ is of strong negative type, therefore, by Theorem 3.16 and Proposition 3.1 in Lyons (2013), there exists an embedding $\phi : \mathbb{R}^\infty \to H$ such that $\|\mathbf{X} - \mathbf{X}^+\| = \|\phi(\mathbf{X}) - \phi(\mathbf{X}^+)\|^2$, and $\alpha_\phi(\mu) = \int \phi(x)\, d\mu(x)$ is injective on the set of signed measures μ on $\mathbb{R}^\infty$ such that $|\mu|$ has a finite first moment i.e., $\int \|x - o\|\, d|\mu|(x) < \infty$, for some $o \in \mathbb{R}^\infty$. Also note

$$
\begin{aligned}
Obj(\boldsymbol{\beta}) &= -E\left[(\varepsilon_t - E(\varepsilon_t))\left(\varepsilon_t^+ - E(\varepsilon_t)\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right] \\
&= -E\left[(\varepsilon_t - E(\varepsilon_t))\left(\varepsilon_t^+ - E(\varepsilon_t)\right) d_\mu\left(\mathbf{X}_t, \mathbf{X}_t^+\right)\right],
\end{aligned}
$$

where $d_\mu\left(\mathbf{X}_t, \mathbf{X}_t^+\right) = \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\| - E_{\mathbf{X}_t^+}\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\| - E_{\mathbf{X}_t}\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\| + E\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|$. Then by Proposition 3.5 in Lyons (2013),

$$
\begin{aligned}
Obj(\boldsymbol{\beta}) &= 2E\left[(\varepsilon_t - E(\varepsilon_t))\left(\varepsilon_t^+ - E(\varepsilon_t)\right)\left\langle \phi(\mathbf{X}_t) - \alpha_\phi(\mu), \phi\left(\mathbf{X}_t^+\right) - \alpha_\phi(\mu)\right\rangle\right] \\
&= 2\left\|E\left[(\varepsilon_t - E(\varepsilon_t))(\phi(\mathbf{X}_t) - \alpha_\phi(\mu))\right]\right\|^2 \geq 0.
\end{aligned}
$$

On the other hand,

$$
Obj(\boldsymbol{\beta}_0) = -E\left[(\varepsilon_{0t} - E(\varepsilon_{0t}))\left(\varepsilon_{0t}^+ - E(\varepsilon_{0t})\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right] = 0,
$$

implying $E\left[(\varepsilon_{0t} - E(\varepsilon_{0t}))(\phi(\mathbf{X}_t) - \alpha_\phi(\mu))\right] = 0$, which can be further simplified as

$$
E\left[\varepsilon_{0t}\phi(\mathbf{X}_t)\right] = 0.
$$

For any Borel set $B \subseteq \mathbb{R}^\infty$, define the signed measure

$$
\mu(B) = E\left[\varepsilon_{0t} 1_B(\mathbf{X}_t)\right].
$$

It is clear that $|\mu(B)|$ has a finite first moment. Then we have

$$\alpha_\phi(\mu(B)) = E[\varepsilon_{0t}\phi(\mathbf{X}_t)] = 0.$$

Then this implies $\mu(B) = 0$ by the injectivity of $\alpha_\phi(\mu)$, i.e.,

$$E[\varepsilon_{0t}1_B(\mathbf{X}_t)] = 0.$$

This implies that

$$E[\varepsilon_{0t}|\mathbf{X}_t] = 0 \text{ a.s..}$$

From another direction, clearly, $E(\varepsilon_{0t}|\mathbf{X}_t) = 0$ a.s. implies $E(\varepsilon_{0t}) = 0$ and $E\left[\varepsilon_{0t}\varepsilon_{0t}^{+}\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right] = 0$, for a fixed q or $q = \infty$. ■

The following Lemmas 9.2 to 9.6 further give some important results regarding integrals involving the non-integrable weighting function, which are useful in the proof of consistency and asymptotic normality of WCIV and WCIVF. Let

$$\hat{\mathbf{Z}}_t(\boldsymbol{\tau}) = \tilde{\mathbf{W}}_t \exp(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle),$$
$$\mathbf{Z}_t(\boldsymbol{\tau}) = (\mathbf{W}_t - \boldsymbol{\mu}_W)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle),$$

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

Lemma 9.2 *Let $\mathbf{W}_t \in \mathbb{R}^p$, $\mathbf{X}_t \in \mathbb{R}^q$. If $(\mathbf{W}_t', \mathbf{X}_t')'$ is i.i.d., and $E\|\mathbf{W}_t\|^2 < \infty$, $E\|\mathbf{X}_t\|^2 < \infty$, then*

$$\int_{\mathbb{R}^q}\frac{1}{n}\sum_{t=1}^{n}\hat{\mathbf{Z}}_t(\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\tau}) \xrightarrow{p} \int_{\mathbb{R}^q}E[\mathbf{Z}_t(\boldsymbol{\tau})]\,\boldsymbol{\omega}(d\boldsymbol{\tau}). \tag{20}$$

Proof. To prove (20), 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

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

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

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

By triangle inequality,

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

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

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

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

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

Similarly, we have

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

We conclude that (20) holds. ■

In Lemma 9.3, 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 $\mathbf{B}_{pn}\left(\boldsymbol{\tau}\right)$ in a Hilbert space. By this approach, the i.i.d. conditions can be relaxed to a weakly stationary time series process conveniently. Specifically, for a fixed δ, $\omega\left(\cdot\right)$ is integrable on $D\left(\delta\right)$, therefore, denote υ as the product measure of $\omega\left(\cdot\right)$ on $D\left(\delta\right)$, i.e., $d\upsilon\left(\boldsymbol{\tau}\right) = \omega\left(d\boldsymbol{\tau}\right)$ on $D\left(\delta\right)$. Then we consider $\mathbf{B}_{pn}\left(\boldsymbol{\tau}\right)$ as a random element in the Hilbert space $L_2\left(D\left(\delta\right),\upsilon\right)$ of all square-integrable q dimensional functions (with respect to the measure υ) with the inner product

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

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

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

An operator $\mathbf{s}$ on a Hilbert space is called nuclear if it can be represented as $\mathbf{s}(\mathbf{h}) = \sum_{j=1}^{\infty} l_j\left\langle\mathbf{h},\mathbf{f}_j\right\rangle_{H(\delta)}\mathbf{f}_j$, 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 $E\|\mathbf{Z}\|^2_{H(\delta)} < \infty$.

Lemma 9.3 *Let $\mathbf{W}_t \in \mathbb{R}^p$, $\mathbf{X}_t \in \mathbb{R}^q$. If $(\mathbf{W}_t', \mathbf{X}_t')'$ is i.i.d., $E(\mathbf{W}_t|\mathbf{X}_t) = \boldsymbol{\mu}_W$, and $E\|\mathbf{W}_t\|^2 < \infty$, $E\|\mathbf{X}_t\|^2 < \infty$, then*

$$\mathbf{B}_{pn}(\boldsymbol{\tau}) \Rightarrow \mathbf{B}_p(\boldsymbol{\tau}), \tag{21}$$

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

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

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

Proof. To prove (21), we show $\mathbf{B}_{pn}(\boldsymbol{\tau})$ is tight by Theorem 2.1 in Politis and Romano (1994). Firstly

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

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 $\|\mathbf{a} + \mathbf{b}\|^2_{H(\delta)} \leq 2\|\mathbf{a}\|^2_{H(\delta)} + 2\|\mathbf{b}\|^2_{H(\delta)}$,

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

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

$$
\begin{aligned}
&\lim_{n\to\infty} E\left\langle \hat{\mathbf{Z}}_1(\boldsymbol{\tau}), \hat{\mathbf{Z}}_K(\boldsymbol{\tau})\right\rangle_{H(\delta)} \\
&= \lim_{n\to\infty} E\left\langle \mathbf{Z}_1(\boldsymbol{\tau}) - (\hat{\boldsymbol{\mu}}_W - \boldsymbol{\mu}_W)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_1\rangle), \mathbf{Z}_K(\boldsymbol{\tau}) - (\hat{\boldsymbol{\mu}}_W - \boldsymbol{\mu}_W)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_K\rangle)\right\rangle_{H(\delta)} \\
&= E\left\langle \mathbf{Z}_1(\boldsymbol{\tau}), \mathbf{Z}_K(\boldsymbol{\tau})\right\rangle_{H(\delta)} - \lim_{n\to\infty} E\left\langle \mathbf{Z}_1(\boldsymbol{\tau}), (\hat{\boldsymbol{\mu}}_W - \boldsymbol{\mu}_W)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_K\rangle)\right\rangle_{H(\delta)} \\
&- \lim_{n\to\infty} E\left\langle \mathbf{Z}_K(\boldsymbol{\tau}), (\hat{\boldsymbol{\mu}}_W - \boldsymbol{\mu}_W)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_1\rangle)\right\rangle_{H(\delta)} \\
&+ \lim_{n\to\infty} E\left\langle (\hat{\boldsymbol{\mu}}_W - \boldsymbol{\mu}_W)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_1\rangle), (\hat{\boldsymbol{\mu}}_W - \boldsymbol{\mu}_W)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_K\rangle)\right\rangle_{H(\delta)} \\
&= 0.
\end{aligned}
$$

Since, for example,

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

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

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

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 Slutskey theorem.

Then the weak convergence follows. Further

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

Then we have

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

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

Lemma 9.4 *Let $\mathbf{W}_t\in\mathbb{R}^p$, $\mathbf{X}_t\in\mathbb{R}^q$. If $(\mathbf{W}_t',\mathbf{X}_t')'$ is i.i.d., and $E\|\mathbf{W}_t\|^2<\infty$, $E\|\mathbf{X}_t\|^2<\infty$. Then*

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

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

Further,

$$\frac{1}{n^2}\sum_{j,k}\tilde{\mathbf{W}}_j'\tilde{\mathbf{W}}_k\left\|X_j-X_k\right\| \xrightarrow{p} E\left[\left(\mathbf{W}_t-\boldsymbol{\mu}_W\right)'\left(\mathbf{W}_t^{+}-\boldsymbol{\mu}_W\right)\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|\right]. \tag{22}$$

If $E\left(\mathbf{W}_t|\mathbf{X}_t\right)=\boldsymbol{\mu}_W$, then

$$\frac{1}{n}\sum_{j,k}\tilde{\mathbf{W}}_j'\tilde{\mathbf{W}}_k\left\|X_j-X_k\right\| = O_p(1). \tag{23}$$

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 9.1. The proof of (22) follows the proof of Theorem 3 in

Shao and Zhang (2014). To prove (23), we need to show

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

For a given δ, by Lemma 9.3 and the continuous mapping theorem, we have

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

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

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

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

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

Therefore

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

Then by Markov's inequality,

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

Finally, by Theorem 8.6.2 of Resnick (1999), we conclude that (23) holds ■

Lemma 9.5 *Let* $\mathbf{W}_t \in \mathbb{R}^p$, $\mathbf{X}_t \in \mathbb{R}^q$. *If* $(\mathbf{W}_t', \mathbf{X}_t')'$ *is i.i.d., and* $E\|\mathbf{W}_t\|^2 < \infty$, $E\|\mathbf{X}_t\|^2 < \infty$. *Then*

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

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

Further,

$$\frac{1}{n^2}\sum_{j,k} \tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_k' \|X_j - X_k\| \xrightarrow{p} E\left[(\mathbf{W}_t - \boldsymbol{\mu}_W)\left(\mathbf{W}_t^+ - \boldsymbol{\mu}_W\right)' \left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right]. \quad (24)$$

If $E(\mathbf{W}_t|\mathbf{X}_t) = \boldsymbol{\mu}_W$, *then*

$$\int_{\mathbb{R}^q} \frac{1}{n} \sum_{t=1}^{n} \hat{\mathbf{Z}}_t(\boldsymbol{\tau}) \left(\sum_{t=1}^{n} \hat{\mathbf{Z}}_t^c(\boldsymbol{\tau}) \right)' \boldsymbol{\omega}(d\boldsymbol{\tau}) = O_p(1). \tag{25}$$

Proof. The analytical forms are proved by repeatedly applying Lemma 9.1. The proof of (24) is analogous to the one for proving (22) in Lemma 9.4. To prove (25), we need to show

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

Again, by Lemma 9.3 and the continuous mapping theorem, for a given $\delta \in (0,1)$, we have

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

When $j = k$, from Lemma 9.4, we have

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

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

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,

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

Similarly we can obtain

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

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

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

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

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

in probability. Then by the continuous mapping theorem,

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

in probability. Finally, by Theorem 8.6.2 of Resnick (1999), we conclude that (25) holds. ■

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

Lemma 9.6 *Let $\mathbf{W}_t \in \mathbb{R}^p$, $\mathbf{X}_t \in \mathbb{R}^q$. If $(\mathbf{W}_t', \mathbf{X}_t')'$ is i.i.d., and $E\|\mathbf{W}_t\|^2 < \infty$, $E\|\mathbf{X}_t\|^2 < \infty$, $E\|\mathbf{f}(\mathbf{X}_t)\|^2 < \infty$. Let $\mathbf{F}_t(\boldsymbol{\tau}) = (\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle)$ and $\hat{\mathbf{F}}_t(\boldsymbol{\tau}) = \tilde{\mathbf{f}}(\mathbf{X}_t)\exp(i\langle\boldsymbol{\tau}, \mathbf{X}_t\rangle)$. Then*

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

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

Further,

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

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

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

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

Lemma 9.7 *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 9.8 *Let* $\tilde{\varepsilon}_{0j}=\varepsilon_{0j}-\frac{1}{n}\sum_{t=1}^{n}\varepsilon_{0t}$, *under Assumptions 1-2,*

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_{0j}D_{jk}\tilde{\varepsilon}_{0k}=o_p(1),$$

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

Proof. Note $E(\varepsilon_{0t}|\mathbf{X}_t)=0$, $D_{jk}=-\|\mathbf{X}_j-\mathbf{X}_k\|$, then by Lemma 9.4,

$$\frac{1}{n}\sum_{j,k}\tilde{\varepsilon}_{0j}D_{jk}\tilde{\varepsilon}_{0k}=O_p(1).$$

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

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_{0j}D_{jk}\tilde{\varepsilon}_{0k}=\frac{1}{r_n^2}O_p(1)=o_p(1).$$

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

As $E(\boldsymbol{\eta}_t|\mathbf{X}_t)=0$, then $\frac{1}{n}\sum_{j,k}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jk}\tilde{\boldsymbol{\eta}}_k'=O_p(1)$, $\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_jD_{jk}\tilde{\mathbf{f}}(\mathbf{X}_k)'=O_p(1)$ by Lemma 9.6; $\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_jD_{jk}\tilde{\boldsymbol{\eta}}_k'=O_p(1)$ by Lemma 9.5. Further $\mathbf{R}_n^{-1}=o_p(1)$ by Assumption 1. So

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

■

Lemma 9.9 *If Assumptions 1-2 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,

$$\begin{aligned}
\frac{1}{nr_n^2}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)'\mathbf{D}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right) &= \frac{1}{nr_n^2}\sum_{j,k}\left(\tilde{y}_j - \tilde{\mathbf{Y}}_j'\boldsymbol{\beta}\right)' D_{jk}\left(\tilde{y}_k - \tilde{\mathbf{Y}}_k'\boldsymbol{\beta}\right) \\
&= \frac{1}{nr_n^2}\sum_{j,k}\left(\tilde{\mathbf{Y}}_j'\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\right) + \tilde{\varepsilon}_{0j}\right)' D_{jk}\left(\tilde{\mathbf{Y}}_k'\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\right) + \tilde{\varepsilon}_{0k}\right) \\
&= \frac{1}{nr_n^2}\left(\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\right)\right)'\left(\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime}\right)\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\right) \\
&+ \frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_{0j}D_{jk}\tilde{\varepsilon}_{0k} + \left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\right)'\frac{2}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_{0k}.
\end{aligned}$$

Note

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_k D_{jk}\tilde{\varepsilon}_{0j} = \frac{1}{nr_n^2}\sum_{j,k}\frac{\mathbf{R}_n\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)}{\sqrt{n}}D_{jk}\tilde{\varepsilon}_{0k} + \frac{1}{nr_n^2}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\varepsilon}_{0k}.$$

Since $E\left(\left(\varepsilon_{0t}, \boldsymbol{\eta}'\right)' | \mathbf{X}_t\right) = 0$, by Lemma 9.5, $\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_j D_{jk}\tilde{\varepsilon}_{0k} = O_p(1)$, by Lemma 9.6, $\frac{1}{n}\sum_{j,k}\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)D_{jk}\tilde{\varepsilon}_{0k} = O_p(1)$. Then we have

$$\frac{1}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_k D_{jk}\tilde{\varepsilon}_{0j} = o_p(1).$$

By Lemma 9.8, $\frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_{0j}D_{jk}\tilde{\varepsilon}_{0k} = o_p(1)$. By Assumption 1, 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 9.8, and $\frac{1}{n}\sum_{j=1}^{n} \tilde{\varepsilon}_{0j}^2 = O_p(1)$, we have

$$\left|\hat{Q}(\boldsymbol{\beta}_0)\right| = \left| \frac{\frac{1}{r_n^2 n}\sum_{j,k} \tilde{\varepsilon}_{0j} D_{jk} \tilde{\varepsilon}_{0k}}{\frac{1}{n}\sum_{j=1}^{n} \tilde{\varepsilon}_{0j}^2} \right| \xrightarrow{p} 0. \tag{27}$$

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 9.7, we arrive at the conclusion. ■

Lemma 9.10 *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

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

$$\begin{aligned}
\frac{1}{nr_n^2}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right)' \mathbf{D} \left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\hat{\boldsymbol{\beta}}\right) &= \frac{1}{nr_n^2}\sum_{j,k}\left(\tilde{y}_j - \tilde{\mathbf{Y}}_j'\hat{\boldsymbol{\beta}}\right)' D_{jk}\left(\tilde{y}_k - \tilde{\mathbf{Y}}_k'\hat{\boldsymbol{\beta}}\right) \\
&= \left(\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}\right)/r_n\right)'\left(\frac{1}{n}\sum_{j,k}\mathbf{R}_n^{-1}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime}\right)\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}\right)/r_n \\
&+ \frac{1}{nr_n^2}\sum_{j,k}\tilde{\varepsilon}_{0j}D_{jk}\tilde{\varepsilon}_{0k} + \left(r_n\mathbf{R}_n'^{-1}\right)\mathbf{R}_n'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}\right)/r_n\frac{2}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_j D_{jk}\tilde{\varepsilon}_{0k}.
\end{aligned}$$

By Lemma 9.8, $\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0$, $\|\mathbf{R}_n^{-1}\| = O\left(r_n^{-1}\right)$, and $\frac{1}{nr_n^2}\sum_{j,k}\tilde{\mathbf{Y}}_k D_{jk}\tilde{\varepsilon}_{0j} = o_p(1)$, 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 $\vartheta_{\min}\left(\mathbf{R}_n\mathbf{R}_n'/r_n^2\right)\geq\vartheta_{\min}\left(\tilde{\mathbf{R}}\tilde{\mathbf{R}}'\right)>0$, 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 9.9. 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

$$\hat{\lambda}_{WCIVF}=o_p\left(r_n^2\right),$$

and

$$\begin{aligned}
&\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}_{WCIVF}-\boldsymbol{\beta}_0\right)/r_n\\
&=\mathbf{R}_n'\left[\tilde{\mathbf{Y}}'\left(\mathbf{D}-\hat{\lambda}_{WCIVF}\mathbf{I}_n\right)\tilde{\mathbf{Y}}\right]^{-1}\tilde{\mathbf{Y}}'\left(\mathbf{D}-\hat{\lambda}_{WCIVF}\mathbf{I}_n\right)\tilde{\boldsymbol{\varepsilon}}_0/r_n\\
&=\left[\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_jD_{jk}\tilde{\mathbf{Y}}_k'-\frac{1}{n}\hat{\lambda}_{WCIVF}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\right)\mathbf{R}_n^{-1\prime}\right]^{-1}\\
&\times\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_jD_{jk}\tilde{\varepsilon}_{0k}-\frac{1}{n}\hat{\lambda}_{WCIVF}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}_0\right)/r_n.
\end{aligned}$$

Since $\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}=O_p\left(n\right)$, $\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}_0=O_p\left(n\right)$, $\|\mathbf{R}_n^{-1}\|=O\left(r_n^{-1}\right)$, therefore

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

$$\mathbf{R}_n^{-1}\frac{1}{n}\hat{\lambda}_{WCIVF}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}_0/r_n=O\left(r_n^{-1}\right)O\left(1/n\right)o_p\left(r_n^2\right)O_p\left(n\right)O\left(r_n^{-1}\right)=o_p\left(1\right).$$

$$\begin{aligned}
\mathbf{R}_n^{-1}\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_jD_{jk}\tilde{\varepsilon}_{0k}/r_n&=\frac{1}{n}\sum_{j,k}\frac{\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)}{\sqrt{n}}D_{jk}\tilde{\varepsilon}_{0k}/r_n+\mathbf{R}_n^{-1}\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_jD_{jk}\tilde{\varepsilon}_{0k}/r_n\\
&=o_p\left(1\right)+o_p\left(1\right)=o_p\left(1\right).
\end{aligned}$$

Further, by Lemma 9.8, 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. For $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_{WCIV}$ or $\hat{\boldsymbol{\beta}}_{WCIVF}$, $\hat{\lambda}=\hat{\lambda}_{WCIV}$ or $\hat{\lambda}_{WCIVF}$,

$$\begin{aligned}\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)&=\mathbf{R}_n'\left[\tilde{\mathbf{Y}}'\left(\mathbf{D}-\hat{\lambda}\mathbf{I}_n\right)\tilde{\mathbf{Y}}\right]^{-1}\tilde{\mathbf{Y}}'\left(\mathbf{D}-\hat{\lambda}\mathbf{I}_n\right)\tilde{\boldsymbol{\varepsilon}}_0\\&=\left[\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_jD_{jk}\tilde{\mathbf{Y}}_k'-\frac{1}{n}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\right)\mathbf{R}_n^{-1\prime}\right]^{-1}\\&\times\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_jD_{jk}\tilde{\varepsilon}_{0k}-\frac{1}{n}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}_0\right),\end{aligned}$$

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

$$\begin{aligned}\mathbf{R}_n^{-1}\frac{1}{\sqrt{n}}\hat{\lambda}\tilde{\mathbf{Y}}'\tilde{\boldsymbol{\varepsilon}}_0&=\mathbf{R}_n^{-1}\frac{1}{n}\hat{\lambda}\sum_j\tilde{\mathbf{Y}}_j\tilde{\varepsilon}_{0j}\\&=\frac{1}{n}\hat{\lambda}\frac{1}{\sqrt{n}}\sum_j\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)\tilde{\varepsilon}_{0k}+\frac{1}{n}\hat{\lambda}\mathbf{R}_n^{-1}\sum_j\tilde{\boldsymbol{\eta}}_j\tilde{\varepsilon}_{0j}\\&=O\left(1/n\right)o_p\left(r_n^2\right)O_p\left(1\right)+O\left(1/\sqrt{n}\right)o_p\left(r_n^2\right)O\left(r_n^{-1}\right)O_p\left(1\right)\\&=o_p\left(1\right),\end{aligned}$$

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

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

By Lemmas 9.8 and 9.5,

$$\mathbf{R}_n^{-1}\left(\frac{1}{n}\sum_{j,k}\tilde{\mathbf{Y}}_jD_{jk}\tilde{\mathbf{Y}}_k'\right)\mathbf{R}_n^{-1\prime}\xrightarrow{p}\boldsymbol{\Pi}.$$

We have

$$
\begin{aligned}
&\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\\
&=\boldsymbol{\Pi}^{-1}\left(\frac{1}{n\sqrt{n}}\sum_{j,k}\tilde{\mathbf{f}}\left(\mathbf{X}_j\right)D_{jk}\tilde{\varepsilon}_{0k}+\frac{\mathbf{R}_n^{-1}}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_jD_{jk}\tilde{\varepsilon}_{0k}\right)+o_p\left(1\right)\\
&=\boldsymbol{\Pi}^{-1}\int_{\mathbb{R}^q}\frac{1}{n}\sum_{j=1}^{n}\hat{\mathbf{F}}_j\left(\boldsymbol{\tau}\right)B_{1n}\left(\boldsymbol{\tau}\right)^c\boldsymbol{\omega}\left(d\boldsymbol{\tau}\right)+o_p\left(1\right)\\
&=:\boldsymbol{\Pi}^{-1}\mathbf{A}_n+o_p\left(1\right),
\end{aligned}
$$

where

$$
B_{1n}\left(\boldsymbol{\tau}\right)=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\tilde{\varepsilon}_{0k}\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_k\right\rangle\right),
$$

because, by Lemma 9.5, $\frac{1}{n}\sum_{j,k}\tilde{\boldsymbol{\eta}}_jD_{jk}\tilde{\varepsilon}_{0k}=O_p\left(1\right)$, $\mathbf{R}_n^{-1}=o\left(1\right)$. By Lemma 9.4

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

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

$$
\begin{aligned}
\Lambda_1\left(\boldsymbol{\tau},\boldsymbol{\varsigma}\right)=&\,E\left[\varepsilon_{0t}^2\exp\left(i\left\langle\boldsymbol{\tau}-\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]+E\left(\varepsilon_{0t}^2\right)E\left[\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]\\
&-E\left[\varepsilon_{0t}^2\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]-E\left[\varepsilon_{0t}^2\exp\left(-i\left\langle\boldsymbol{\varsigma},\mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right],
\end{aligned}
$$

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

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

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

Then in the same spirit of proving Lemma 9.5, by Cauchy-Schwarz inequality

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

Then by the dominated convergence theorem,

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

Similarly we can obtain

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

so we conclude that

$$
\mathbf{A}_n \xrightarrow{d} \int_{\mathbb{R}^q} E\left[\mathbf{F}_t(\boldsymbol{\tau})\right] B_1(\boldsymbol{\tau})^c \boldsymbol{\omega}(d\boldsymbol{\tau}),
$$

where the integrated weighted Gaussian process follows a normal distribution with mean zero and variance

$$
\mathbf{V}(\boldsymbol{\theta}_0) = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\mathbf{F}_t(\boldsymbol{\tau})\right] E\left[\mathbf{F}_t(-\boldsymbol{\varsigma})\right]' \Lambda_1(\boldsymbol{\tau},\boldsymbol{\varsigma})^c \boldsymbol{\omega}(d\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\varsigma}).
$$

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

$$
\begin{aligned}
\mathbf{V}(\boldsymbol{\theta}_0) = & \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\mathbf{F}_t(\boldsymbol{\tau})\right] E\left[\mathbf{F}_t(-\boldsymbol{\varsigma})\right]' \\
& \times\left\{\begin{array}{c}
E\left[\varepsilon_{0t}^2 \exp\left(-i\left\langle\boldsymbol{\tau}-\boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right] + E\left(\varepsilon_{0t}^2\right) E\left[\exp\left(-i\left\langle\boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right] E\left[\exp\left(i\left\langle\boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right] \\
-E\left[\varepsilon_{0t}^2 \exp\left(-i\left\langle\boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right] E\left[\exp\left(i\left\langle\boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right] - E\left[\varepsilon_{0t}^2 \exp\left(i\left\langle\boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right] E\left[\exp\left(-i\left\langle\boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right]
\end{array}\right\} \boldsymbol{\omega}(d\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\varsigma} \\
=: & \mathbf{V}_1(\boldsymbol{\theta}_0) + \mathbf{V}_2(\boldsymbol{\theta}_0) - \mathbf{V}_3(\boldsymbol{\theta}_0) - \mathbf{V}_4(\boldsymbol{\theta}_0).
\end{aligned}
$$

By Fubini's Theorem and Lemma 9.1,

$$\mathbf{V}_1(\boldsymbol{\theta}_0) = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\begin{array}{c} \left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t^+\right\rangle\right)\left(\mathbf{f}\left(\mathbf{X}_t^{++}\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\exp\left(-i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^{++}\right\rangle\right) \\ \times\varepsilon_{0t}^2\exp\left(-i\left\langle \boldsymbol{\tau}-\boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right) \end{array}\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\begin{array}{c} \varepsilon_{0t}^2\left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t^+ - \mathbf{X}_t\right\rangle\right) \\ \times\left(\mathbf{f}\left(\mathbf{X}_t^{++}\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)'\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t - \mathbf{X}_t^{++}\right\rangle\right) \end{array}\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\begin{array}{c} \varepsilon_{0t}^2\left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\left(1-\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right) \\ \times\left(\mathbf{f}\left(\mathbf{X}_t^{++}\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)'\left(1-\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t - \mathbf{X}_t^{++}\right\rangle\right)\right) \end{array}\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= E\left[\begin{array}{c} \int_{\mathbb{R}^q}\varepsilon_{0t}^2\left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\left[1-\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau}) \\ \times\int_{\mathbb{R}^q}\left(\mathbf{f}\left(\mathbf{X}_t^{++}\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)'\left[1-\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t - \mathbf{X}_t^{++}\right\rangle\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma}) \end{array}\right]$$

$$= E\left[\varepsilon_{0t}^2\left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\left(\mathbf{f}\left(\mathbf{X}_t^{++}\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)'\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\left\|\mathbf{X}_t - \mathbf{X}_t^{++}\right\|\right].$$

$$\mathbf{V}_2(\boldsymbol{\theta}_0) = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q}\left\{\begin{array}{c} E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right]E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\exp\left(-i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right]' \\ \times E\left(\varepsilon_{0t}^2\right)E\left[\exp\left(-i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right] \end{array}\right\}\boldsymbol{\omega}(d\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= E\left(\varepsilon_t^2\right)\int_{\mathbb{R}^q} E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\left(1-\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau})$$

$$\times\int_{\mathbb{R}^q} E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})'\left(1-\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^+ - \mathbf{X}_t\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= E\left(\varepsilon_{0t}^2\right)E\left((\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right)E\left((\mathbf{f}(\mathbf{X}_t) - E\mathbf{f}(\mathbf{X}_t))'\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right).$$

$$\mathbf{V}_3(\boldsymbol{\theta}_0) = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q}\left\{\begin{array}{c} E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right]E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\exp\left(-i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right]' \\ \times E\left[\varepsilon_{0t}^2\exp\left(-i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right]E\left[\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t\right\rangle\right)\right] \end{array}\right\}\boldsymbol{\omega}(d\boldsymbol{\tau})\,\boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= \int_{\mathbb{R}^q} E\left[\varepsilon_{0t}^2\left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\left(1-\exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\tau})$$

$$\times\int_{\mathbb{R}^q} E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})'\left(1-\exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^+ - \mathbf{X}_t\right\rangle\right)\right)\right]\boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= E\left(\varepsilon_{0t}^2\left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right)E\left((\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})'\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right).$$

$$\mathbf{V}_4(\boldsymbol{\theta}_0) = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} \left\{ \begin{array}{c} E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}}) \exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right] E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}}) \exp(-i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right]' \\ E\left[\varepsilon_{0t}^2 \exp(i\langle \boldsymbol{\varsigma}, \mathbf{X}_t\rangle)\right] E\left[\exp(-i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right] \end{array} \right\} \boldsymbol{\omega}(d\boldsymbol{\tau})\, \boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= \int_{\mathbb{R}^q} E\left[(\mathbf{f}(\mathbf{X}_t) - \boldsymbol{\mu}_{\mathbf{f}})\left(1 - \exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\right] \boldsymbol{\omega}(d\boldsymbol{\tau})$$

$$\times \int_{\mathbb{R}^q} E\left[\varepsilon_{0t}^2 \left(\mathbf{f}\left(\mathbf{X}_t^+\right) - \boldsymbol{\mu}_{\mathbf{f}}\right)' \left(1 - \exp\left(i\left\langle \boldsymbol{\varsigma}, \mathbf{X}_t^+ - \mathbf{X}_t\right\rangle\right)\right)\right] \boldsymbol{\omega}(d\boldsymbol{\varsigma})$$

$$= \mathbf{V}_3(\boldsymbol{\theta}_0)'.$$

■

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

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

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

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

since by Lemma 9.10, $\hat{\lambda}/n = o_p\left(r_n^2/n\right) = o_p(1)$, and

$$\frac{1}{n}\mathbf{R}_n^{-1}\sum_{l=1}^{n}\sum_{j=1}^{n} \tilde{\mathbf{Y}}_j \tilde{\mathbf{Y}}_l' \varepsilon_l\left(\hat{\boldsymbol{\theta}}\right)^2 D_{jl}\mathbf{R}_n^{-1\prime} = O_p(1).$$

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

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

Now

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

$$
\begin{aligned}
\frac{\mathbf{R}_n^{-1}}{\sqrt{n}}\sum_{j=1}^{n}\tilde{\mathbf{Y}}_jD_{jl} &= \frac{1}{n}\sum_{j=1}^{n}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jl} + \frac{\mathbf{R}_n^{-1}}{\sqrt{n}}\sum_{j=1}^{n}\tilde{\boldsymbol{\eta}}_jD_{jl} \\
&= -E_j\left[(\mathbf{f}(\mathbf{X}_j) - \boldsymbol{\mu}_{\mathbf{f}})\|\mathbf{X}_j - \mathbf{X}_l\|\right] + o_p(1),
\end{aligned}
$$

since

$$
\frac{\mathbf{R}_n^{-1}}{\sqrt{n}}\sum_{j=1}^{n}\tilde{\boldsymbol{\eta}}_jD_{jl} = o_p(1),
$$

$$
\frac{1}{n}\sum_{j=1}^{n}\tilde{\mathbf{f}}(\mathbf{X}_j)D_{jl} \xrightarrow{p} -E_j\left[(\mathbf{f}(\mathbf{X}_j) - \boldsymbol{\mu}_{\mathbf{f}})\|\mathbf{X}_j - \mathbf{X}_l\|\right],
$$

where E_j denotes the expectation in terms of $(\mathbf{Y}_j, \mathbf{X}_j)$. Similarly,

$$
\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime}D_{kl} = -E_k\left[(\mathbf{f}(\mathbf{X}_k) - \boldsymbol{\mu}_{\mathbf{f}})'\|\mathbf{X}_k - \mathbf{X}_l\|\right] + o_p(1).
$$

So by the continuous mapping theorem, we conclude that

$$
n\mathbf{R}_n^{-1}\hat{\boldsymbol{\Omega}}_1\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)\mathbf{R}_n^{-1\prime} \xrightarrow{p} \mathbf{V}_1(\boldsymbol{\theta}_0).
$$

Analogously we can show

$$
n\mathbf{R}_n^{-1}\hat{\boldsymbol{\Omega}}_2\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)\mathbf{R}_n^{-1\prime} \xrightarrow{p} \mathbf{V}_2(\boldsymbol{\theta}_0),
$$

$$
n\mathbf{R}_n^{-1}\hat{\boldsymbol{\Omega}}_3\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)\mathbf{R}_n^{-1\prime} \xrightarrow{p} \mathbf{V}_3(\boldsymbol{\theta}_0).
$$

Then by the continuous mapping theorem,

$$
n\mathbf{R}_n^{-1}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)\mathbf{R}_n^{-1\prime} \xrightarrow{p} \mathbf{V}(\boldsymbol{\theta}_0).
$$

Next

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

Since

$$\frac{\mathbf{R}_n^{-1}}{n}\sum_{j,k}\tilde{\mathbf{Y}}_jD_{jk}\tilde{\mathbf{Y}}_k'\mathbf{R}_n^{-1\prime}=\boldsymbol{\Pi}+o_p(1)$$

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

we have

$$n\mathbf{R}_n^{-1}\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1\prime}\xrightarrow{p}\boldsymbol{\Pi}.$$

Therefore,

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

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

$$\frac{1}{n}\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\theta}},\hat{\lambda}\right)\hat{\boldsymbol{\Upsilon}}\left(\hat{\lambda}\right)^{-1}\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)'$$

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

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

■

9.1 Discussion on the Efficiency of WCIV and WCIVF

We analyze the efficiency of WCIV and WCIVF for fixed $q=1,2$ in an i.i.d. setup under homoskedasticity and two distributional assumptions on $(\mathbf{Y},\mathbf{X})$. The matrix $\boldsymbol{\Upsilon}$ of squared expected derivatives of the moment conditions corresponding to the objective function

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

for

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

and $\tilde{y}_t = y_t - \bar{y}$, evaluated at the true parameter value $\boldsymbol{\beta}_0$ can be represented as

$$\begin{aligned}\boldsymbol{\Upsilon} &= \int E\left[\frac{\partial}{\partial\boldsymbol{\beta}}h\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)\right] E\left[\frac{\partial}{\partial\boldsymbol{\beta}}h\left(\boldsymbol{\beta}_0,-\boldsymbol{\tau}\right)\right]' \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \\ &= \int E\left[\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right] E\left[\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\exp\left(-i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right]' \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right),\end{aligned}$$

where for $\mathbf{Z}_t = \left((\mathbf{Y_t} - \boldsymbol{\mu}_\mathbf{Y})', \mathbf{X}_t'\right)'$,

$$E\left[\frac{\partial}{\partial\boldsymbol{\beta}}h\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)\right] = -E\left[\left(\mathbf{Y_t} - \boldsymbol{\mu}_\mathbf{Y}\right)\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right] = -\frac{1}{i}\frac{\partial}{\partial\boldsymbol{\lambda}}\varphi_\mathbf{Z}\left(\boldsymbol{\lambda},\boldsymbol{\tau}\right)\bigg|_{\boldsymbol{\lambda}=\mathbf{0}}.$$

The asymptotic variance of the normalized score evaluated at the true value $\boldsymbol{\beta}_0$,

$$n^{1/2}\int_{\mathbb{R}^q}\frac{\partial}{\partial\boldsymbol{\beta}}h_n\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)h_n\left(\boldsymbol{\beta}_0,-\boldsymbol{\tau}\right)\boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \to_d n^{1/2}\,\mathrm{Re}\int_{\mathbb{R}^q} E\left[\frac{\partial}{\partial\boldsymbol{\beta}}h\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)\right]h_n\left(\boldsymbol{\beta}_0,-\boldsymbol{\tau}\right)\boldsymbol{\omega}\left(d\boldsymbol{\tau}\right),$$

is

$$\boldsymbol{\Omega} = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} E\left[\frac{\partial}{\partial\boldsymbol{\beta}}h\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)\right] E\left[\frac{\partial}{\partial\boldsymbol{\beta}}h\left(\boldsymbol{\beta}_0,-\boldsymbol{\lambda}\right)\right]' k\left(\boldsymbol{\tau},\boldsymbol{\lambda}\right)\boldsymbol{\omega}\left(d\boldsymbol{\tau}\right)\boldsymbol{\omega}\left(d\boldsymbol{\lambda}\right)$$

where we set

$$\begin{aligned}k\left(\boldsymbol{\tau},\boldsymbol{\lambda}\right) &:= \lim_{n\to\infty} E\left[n h_n\left(\boldsymbol{\beta}_0,\boldsymbol{\tau}\right)h_n\left(\boldsymbol{\beta}_0,-\boldsymbol{\lambda}\right)\right] \\ &= E\left[\varepsilon_{0t}^2\left(\begin{array}{c}\exp\left(i\left\langle\boldsymbol{\tau}-\boldsymbol{\lambda},\mathbf{X}_t\right\rangle\right) - \exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle - i\left\langle\boldsymbol{\lambda},\mathbf{X}_t^+\right\rangle\right) \\ -\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t^+\right\rangle - i\left\langle\boldsymbol{\lambda},\mathbf{X}_t\right\rangle\right) + \varphi_\mathbf{X}\left(\boldsymbol{\tau}\right)\varphi_\mathbf{X}\left(-\boldsymbol{\lambda}\right)\end{array}\right)\right],\end{aligned}$$

and under homoskedasticity, $\sigma_\varepsilon^2 = E\left[\varepsilon_{0t}^2\right] = E\left[\varepsilon_{0t}^2|\mathbf{X}_t\right]$,

$$\begin{aligned}k\left(\boldsymbol{\tau},\boldsymbol{\lambda}\right) &= \sigma_\varepsilon^2 E\left[\begin{array}{c}\exp\left(i\left\langle\boldsymbol{\tau}-\boldsymbol{\lambda},\mathbf{X}_t\right\rangle\right) - \exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle - i\left\langle\boldsymbol{\lambda},\mathbf{X}_t^+\right\rangle\right) \\ -\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t^+\right\rangle - i\left\langle\boldsymbol{\lambda},\mathbf{X}_t\right\rangle\right) + \varphi_\mathbf{X}\left(\boldsymbol{\tau}\right)\varphi_\mathbf{X}\left(-\boldsymbol{\lambda}\right)\end{array}\right] \\ &= \sigma_\varepsilon^2\left[\varphi_\mathbf{X}\left(\boldsymbol{\tau}-\boldsymbol{\lambda}\right) - \varphi_\mathbf{X}\left(\boldsymbol{\tau}\right)\varphi_\mathbf{X}\left(-\boldsymbol{\lambda}\right)\right].\end{aligned}$$

9.1.1 Normal case with non-integrable kernel

Assuming $\mathbf{Z}_t \sim N(\mu_{\mathbf{Z}}, \Sigma_{\mathbf{ZZ}})$ for

$$\mu_{\mathbf{Z}} = \begin{pmatrix} 0 \\ \boldsymbol{\mu}_X \end{pmatrix}, \quad \Sigma_{\mathbf{ZZ}} = \begin{pmatrix} \Sigma_{\mathbf{YY}} & \Sigma_{\mathbf{YX}} \\ \Sigma_{\mathbf{XY}} & \Sigma_{\mathbf{XX}} \end{pmatrix}$$

we have

$$\varphi_{\mathbf{Z}}(\boldsymbol{\lambda}, \boldsymbol{\tau}) = \exp\left(i \langle \boldsymbol{\mu}_X, \boldsymbol{\tau} \rangle - \frac{1}{2}(\boldsymbol{\lambda}', \boldsymbol{\tau}') \Sigma_{\mathbf{ZZ}} (\boldsymbol{\lambda}', \boldsymbol{\tau}')'\right)$$

and

$$\left.\frac{\partial}{\partial \boldsymbol{\lambda}} \varphi_{\mathbf{Z}}(\boldsymbol{\lambda}, \boldsymbol{\tau})\right|_{\boldsymbol{\lambda}=\mathbf{0}} = -\Sigma_{\mathbf{YX}} \boldsymbol{\tau} \varphi_{\mathbf{Z}}(\mathbf{0}, \boldsymbol{\tau}) = -\Sigma_{\mathbf{YX}} \boldsymbol{\tau} \varphi_{\mathbf{X}}(\boldsymbol{\tau}).$$

Then,

$$\begin{aligned} \boldsymbol{\Upsilon} &= \int_{\mathbb{R}^q} \Sigma_{\mathbf{YX}} \boldsymbol{\tau}\boldsymbol{\tau}' \Sigma_{\mathbf{XY}} \left|\varphi_{\mathbf{X}}(\boldsymbol{\tau})\right|^2 \boldsymbol{\omega}(d\boldsymbol{\tau}) \\ &= \Sigma_{\mathbf{YX}} \left\{ \int_{\mathbb{R}^q} \boldsymbol{\tau}\boldsymbol{\tau}' \exp\left(-\boldsymbol{\tau}' \Sigma_{\mathbf{XX}} \boldsymbol{\tau}\right) \boldsymbol{\omega}(d\boldsymbol{\tau}) \right\} \Sigma_{\mathbf{XY}} \\ &= \Sigma_{\mathbf{YX}} \left\{ \int_{\mathbb{R}^q} \frac{\boldsymbol{\tau}\boldsymbol{\tau}' \exp\left(-\boldsymbol{\tau}' \Sigma_{\mathbf{XX}} \boldsymbol{\tau}\right)}{c_q \left\|\boldsymbol{\tau}\right\|^{q+1}} d\boldsymbol{\tau} \right\} \Sigma_{\mathbf{XY}}, \end{aligned}$$

where the term in braces is bounded. Assuming homoskedastic and independent elements in $\mathbf{X}$, $\Sigma_{\mathbf{XX}} = \sigma_{\mathbf{X}}^2 \mathbf{I}_q$,

$$\int_{\mathbb{R}^q} \frac{\boldsymbol{\tau}\boldsymbol{\tau}' \exp\left(-\boldsymbol{\tau}' \Sigma_{\mathbf{XX}} \boldsymbol{\tau}\right)}{c_q \left\|\boldsymbol{\tau}\right\|^{q+1}} d\boldsymbol{\tau} = \int_{\mathbb{R}^q} \frac{\boldsymbol{\tau}\boldsymbol{\tau}' \exp\left(-\sigma_X^2 \boldsymbol{\tau}' \boldsymbol{\tau}\right)}{c_q \left\|\boldsymbol{\tau}\right\|^{q+1}} d\boldsymbol{\tau}.$$

When $q = 1$, for $\rho_{\mathbf{YX}} = \sigma_{\mathbf{YX}} / (\sigma_{\mathbf{Y}} \sigma_{\mathbf{X}})$,

$$\begin{aligned} \boldsymbol{\Upsilon} &= \sigma_{\mathbf{YX}}^2 \int_{\mathbb{R}} \boldsymbol{\tau}^2 \exp\left(-\boldsymbol{\tau}^2 \sigma_{\mathbf{X}}^2\right) \frac{1}{\pi \boldsymbol{\tau}^2} d\boldsymbol{\tau} \\ &= \sigma_{\mathbf{YX}}^2 \frac{\left(2\pi \left(\sigma_{\mathbf{X}}^{-2}/2\right)\right)^{1/2}}{\pi} \frac{1}{\left(2\pi \left(\sigma_{\mathbf{X}}^{-2}/2\right)\right)^{1/2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2\left(\sigma_{\mathbf{X}}^{-2}/2\right)} \boldsymbol{\tau}^2\right) d\boldsymbol{\tau} \\ &= \sigma_{\mathbf{YX}}^2 \frac{\left(2\pi \left(\sigma_{\mathbf{X}}^{-2}/2\right)\right)^{1/2}}{\pi} = \frac{\sigma_{\mathbf{YX}}^2}{\sqrt{\pi} \sigma_{\mathbf{X}}} = \frac{1}{\sqrt{\pi}} \rho_{\mathbf{YX}}^2 \sigma_{\mathbf{Y}}^2 \sigma_{\mathbf{X}}. \end{aligned}$$

When $q=2$, for $i=1,2$,

$$\begin{aligned}
\int_{\mathbb{R}^2} \frac{\boldsymbol{\tau}_i^2 \exp\left(-\sigma_X^2 \boldsymbol{\tau}'\boldsymbol{\tau}\right)}{c_q \|\boldsymbol{\tau}\|^{q+1}} d\boldsymbol{\tau} &= \frac{1}{c_2}\int_{\mathbb{R}^2} \frac{\boldsymbol{\tau}_1^2 \exp\left(-\sigma_X^2\left(\boldsymbol{\tau}_1^2+\boldsymbol{\tau}_2^2\right)\right)}{\left(\boldsymbol{\tau}_1^2+\boldsymbol{\tau}_2^2\right)^{3/2}} d\boldsymbol{\tau}_1 d\boldsymbol{\tau}_2 \\
&= \frac{1}{2c_2}\lim_{r\to\infty}\int_{-r}^{r}\int_0^{2\pi} s^2\cos^2(\theta)\frac{\exp\left(-\sigma_X^2 s^2\right)}{s^3} s ds d\theta \\
&= \frac{1}{2c_2}\lim_{r\to\infty}\int_{-r}^{r}\int_0^{2\pi} \cos^2(\theta)\exp\left(-\sigma_X^2 s^2\right) ds d\theta \\
&= \frac{1}{2c_2}(\pi)^{3/2}/\sigma_{\mathbf{X}} = \frac{\sqrt{\pi}}{4\sigma_{\mathbf{X}}},
\end{aligned}$$

using

$$\begin{aligned}
\boldsymbol{\tau}_1 &= s\cos(\theta), \boldsymbol{\tau}_2 = s\sin(\theta) \\
d\boldsymbol{\tau}_1 d\boldsymbol{\tau}_2 &= s ds d\theta \\
\int_0^{2\pi}\cos^2(\theta) d\theta &= \frac{1}{2}\sin(\theta)\cos(\theta)|_0^{2\pi} + \frac{1}{2}\theta|_0^{2\pi} = \pi \\
\lim_{r\to\infty}\int_{-r}^{r}\exp\left(-\sigma_X^2 s^2\right) ds &= (\pi)^{1/2}/\sigma_{\mathbf{X}},
\end{aligned}$$

while, proceeding similarly,

$$\begin{aligned}
\int_{\mathbb{R}^2} \frac{\boldsymbol{\tau}_1\boldsymbol{\tau}_2 \exp\left(-\sigma_X^2 \boldsymbol{\tau}'\boldsymbol{\tau}\right)}{c_2 \|\boldsymbol{\tau}\|^{q+1}} d\boldsymbol{\tau} &= \frac{\Gamma(3/2)}{\pi^{3/2}}\int_{\mathbb{R}^2} \frac{\boldsymbol{\tau}_1\boldsymbol{\tau}_2 \exp\left(-\sigma_X^2\left(\boldsymbol{\tau}_1^2+\boldsymbol{\tau}_2^2\right)\right)}{\|\boldsymbol{\tau}_1^2+\boldsymbol{\tau}_2^2\|^{3/2}} d\boldsymbol{\tau}_1 d\boldsymbol{\tau}_2 \\
&= \frac{1}{2c_2}\lim_{r\to\infty}\int_{-r}^{r}\int_0^{2\pi} s^2\cos(\theta)\sin(\theta)\frac{\exp\left(-\sigma_X^2 s^2\right)}{s^3} s ds d\theta \\
&= \frac{1}{2c_2}\int_{-r}^{r}\int_0^{2\pi}\cos(\theta)\sin(\theta)\exp\left(-\sigma_X^2 s^2\right) ds d\theta \\
&= 0,
\end{aligned}$$

using

$$\int_0^{2\pi}\cos(\theta)\sin(\theta) d\theta = \frac{1}{2}\sin^2(\theta)|_0^{2\pi} = 0,$$

so that when $q=2$

$$\boldsymbol{\Upsilon} = \frac{\sqrt{\pi}}{4\sigma_{\mathbf{X}}}\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XY}}.$$

Next,

$$\boldsymbol{\Omega} = \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} \Sigma_{\mathbf{YX}}\boldsymbol{\tau}\boldsymbol{\lambda}'\Sigma_{\mathbf{XY}}\varphi_{\mathbf{X}}(\boldsymbol{\tau})\varphi_{\mathbf{X}}(-\boldsymbol{\lambda}) k(\boldsymbol{\lambda},\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\lambda}) = \sigma_\varepsilon^2\Sigma_{\mathbf{YX}}\mathbf{A}\Sigma_{\mathbf{XY}}$$

where

$$\begin{aligned}\mathbf{A} &:= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} \boldsymbol{\tau}\boldsymbol{\lambda}'\varphi_{\mathbf{X}}(\boldsymbol{\tau})\varphi_{\mathbf{X}}(-\boldsymbol{\lambda})\left[\varphi_{\mathbf{X}}(\boldsymbol{\lambda}-\boldsymbol{\tau})-\varphi_{\mathbf{X}}(\boldsymbol{\lambda})\varphi_{\mathbf{X}}(-\boldsymbol{\tau})\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\lambda}) \\ &= \int_{\mathbb{R}^q}\int_{\mathbb{R}^q} \boldsymbol{\tau}\boldsymbol{\lambda}'\left[\varphi_{\mathbf{X}}(\boldsymbol{\tau})\varphi_{\mathbf{X}}(-\boldsymbol{\lambda})\varphi_{\mathbf{X}}(\boldsymbol{\lambda}-\boldsymbol{\tau})-|\varphi_{\mathbf{X}}(\boldsymbol{\tau})|^2|\varphi_{\mathbf{X}}(\boldsymbol{\lambda})|^2\right]\boldsymbol{\omega}(d\boldsymbol{\tau})\boldsymbol{\omega}(d\boldsymbol{\lambda}).\end{aligned}$$

When $q = 1$,

$$\begin{aligned}\mathbf{A} &= \int_{\mathbb{R}^2} \frac{\tau\lambda}{\pi^2\tau^2\lambda^2}\left[\varphi_{\mathbf{X}}(\boldsymbol{\tau})\varphi_{\mathbf{X}}(-\boldsymbol{\lambda})\varphi_{\mathbf{X}}(\boldsymbol{\lambda}-\boldsymbol{\tau})-|\varphi_{\mathbf{X}}(\boldsymbol{\tau})|^2|\varphi_{\mathbf{X}}(\boldsymbol{\lambda})|^2\right]d\tau d\lambda \\ &= \int_{\mathbb{R}^2} \frac{1}{\pi^2\tau\lambda}\left[\exp\begin{pmatrix} -\frac{1}{2}\tau^2\sigma_{\mathbf{X}}^2-\frac{1}{2}\lambda^2\sigma_{\mathbf{X}}^2 \\ -\frac{1}{2}(\tau-\lambda)^2\sigma_{\mathbf{X}}^2\end{pmatrix} - \exp\left(-\tau^2\sigma_{\mathbf{X}}^2-\lambda^2\sigma_{\mathbf{X}}^2\right)\right]d\tau d\lambda \\ &= \int_{\mathbb{R}^2} \frac{1}{\pi^2\tau\lambda}\exp\left(-\tau^2\sigma_{\mathbf{X}}^2-\lambda^2\sigma_{\mathbf{X}}^2\right)\left[\exp\left(\tau\lambda\sigma_{\mathbf{X}}^2\right)-1\right]d\tau d\lambda \\ &= \sum_{j=1}^{\infty}\int_{\mathbb{R}^2}\frac{1}{\pi^2\tau\lambda}\exp\left(-\tau^2\sigma_{\mathbf{X}}^2-\lambda^2\sigma_{\mathbf{X}}^2\right)\frac{\left(\tau\lambda\sigma_{\mathbf{X}}^2\right)^{2j-1}}{(2j-1)!}d\tau d\lambda \\ &= 4\sum_{j=0}^{\infty}\frac{\left(\sigma_{\mathbf{X}}^2\right)^{2j+1}}{(2j+1)!}\left(\frac{1}{2\pi}\int_{\mathbb{R}}\exp\left(-\frac{1}{2}\tau^2 2\sigma_{\mathbf{X}}^2\right)\tau^{2j}d\tau\right)^2 \\ &= 4\sum_{j=0}^{\infty}\frac{\left(\sigma_{\mathbf{X}}^2\right)^{2j+1}}{(2j+1)!}\left(\frac{\left(2^{-j}\frac{(2j)!}{j!}\right)^2}{2\pi}\left\{2\sigma_{\mathbf{X}}^2\right\}^{-(2j+1)}\right) \\ &= \frac{1}{\pi}\sum_{j=0}^{\infty}\frac{1}{(2j+1)!}\left(\frac{(2j)!}{j!}\right)^2 2^{-4j} = \frac{1}{3},\end{aligned}$$

so that

$$\boldsymbol{\Omega} = \frac{1}{3}\sigma_{\varepsilon}^2\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XY}} = \frac{1}{3}\sigma_{\varepsilon}^2\rho_{\mathbf{YX}}^2\sigma_{\mathbf{X}}^2\sigma_{\mathbf{Y}}^2$$

and the asymptotic variance of $WCIV$ for $q = 1$ is

$$\begin{aligned}\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{-1} &= \left(\frac{1}{\sqrt{\pi}}\rho_{\mathbf{YX}}^2\sigma_{\mathbf{Y}}^2\sigma_{\mathbf{X}}\right)^{-2}\frac{1}{3}\sigma_{\varepsilon}^2\rho_{\mathbf{YX}}^2\sigma_{\mathbf{X}}^2\sigma_{\mathbf{Y}}^2 \\ &= \frac{\pi}{3}\frac{\sigma_{\varepsilon}^2}{\rho_{\mathbf{YX}}^2\sigma_{\mathbf{Y}}^2} \approx 1.0472\frac{\sigma_{\varepsilon}^2}{\rho_{\mathbf{YX}}^2\sigma_{\mathbf{Y}}^2},\end{aligned}$$

which is invariant to $\sigma_{\mathbf{X}}^2$ and only marginally larger than the limiting optimal Gaussian-weighting case and the IV case, $\sigma_{\varepsilon}^2/\left(\rho_{\mathbf{YX}}^2\sigma_{\mathbf{Y}}^2\right)$ with $ARE_{WCIV} = \pi/3 \approx 1.0472$.

When $q = 2$ or larger we have

$$\mathbf{A} = \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} \frac{\boldsymbol{\tau}\boldsymbol{\lambda}'}{c_q^2 \|\boldsymbol{\tau}\|^{q+1} \|\boldsymbol{\lambda}\|^{q+1}} \exp\left(-\sigma_X^2 \|\boldsymbol{\tau}\|^2 - \sigma_X^2 \|\boldsymbol{\lambda}\|^2\right) \left[\exp\left(\sigma_X^2 \boldsymbol{\tau}'\boldsymbol{\lambda}\right) - 1\right] d\boldsymbol{\tau} d\boldsymbol{\lambda}$$

where

$$\boldsymbol{\tau}\boldsymbol{\lambda}' \left[\exp\left(\sigma_X^2 \boldsymbol{\tau}'\boldsymbol{\lambda}\right) - 1\right] = \boldsymbol{\tau}\boldsymbol{\lambda}' \sum_{j=1}^{\infty} \frac{\left(\sigma_X^2 \boldsymbol{\tau}'\boldsymbol{\lambda}\right)^j}{j!} = \boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\lambda}\boldsymbol{\lambda}' \sum_{j=1}^{\infty} \frac{\sigma_X^{2j} \left(\boldsymbol{\tau}'\boldsymbol{\lambda}\right)^{j-1}}{j!}$$

and for $q = 2$ and $j = 1, 2, \ldots,$

$$\begin{aligned} \left(\boldsymbol{\tau}'\boldsymbol{\lambda}\right)^{j-1} &= \left(\tau_1\lambda_1 + \tau_2\lambda_2\right)^{j-1} = \sum_{\ell=0}^{j-1} \tbinom{j-1}{\ell} \left(\tau_1\lambda_1\right)^{\ell} \left(\tau_2\lambda_2\right)^{j-1-\ell} \\ &= \sum_{\ell=0}^{j-1} \tbinom{j-1}{\ell} \left(\tau_1\right)^{\ell} \left(\tau_2\right)^{j-1-\ell} \left(\lambda_1\right)^{\ell} \left(\lambda_2\right)^{j-1-\ell}. \end{aligned}$$

Then

$$\begin{aligned} \mathbf{A} &= \sum_{j=1}^{\infty} \frac{\sigma_X^{2j}}{j!} \sum_{\ell=0}^{j-1} \tbinom{j-1}{\ell} \int_{\mathbb{R}^q} \frac{\boldsymbol{\tau}\boldsymbol{\tau}' \exp\left(-\sigma_X^2 \|\boldsymbol{\tau}\|^2\right)}{c_q \|\boldsymbol{\tau}\|^{q+1}} \left(\tau_1\right)^{\ell} \left(\tau_2\right)^{j-1-\ell} d\boldsymbol{\tau} \\ &\quad \times \int_{\mathbb{R}^q} \frac{\boldsymbol{\lambda}\boldsymbol{\lambda}' \exp\left(-\sigma_X^2 \|\boldsymbol{\lambda}\|^2\right)}{c_q \|\boldsymbol{\lambda}\|^{q+1}} \left(\lambda_1\right)^{\ell} \left(\lambda_2\right)^{j-1-\ell} d\boldsymbol{\lambda} \\ &= \sum_{j=1}^{\infty} \frac{\sigma_X^{2j}}{j!} \sum_{\ell=0}^{j-1} \tbinom{j-1}{\ell} A_{j,\ell}^2, \quad A_{j,\ell} = \int_{\mathbb{R}^q} \frac{\boldsymbol{\tau}\boldsymbol{\tau}' \exp\left(-\sigma_X^2 \|\boldsymbol{\tau}\|^2\right)}{c_q \|\boldsymbol{\tau}\|^{q+1}} \left(\tau_1\right)^{\ell} \left(\tau_2\right)^{j-1-\ell} d\boldsymbol{\tau}, \end{aligned}$$

where we evaluate the typical elements in the matrix $A_{j,\ell}$, diagonal and off-diagonal,

$$\text{diagonal:} \qquad a_{j,\ell} := \int_{\mathbb{R}^q} \frac{\tau_1^2 \exp\left(-\sigma_X^2 \|\boldsymbol{\tau}\|^2\right)}{c_q \|\boldsymbol{\tau}\|^{q+1}} \left(\tau_1\right)^{\ell} \left(\tau_2\right)^{j-1-\ell} d\boldsymbol{\tau}, \quad \ell \textit{ even}$$

$$\text{off-diagonal:} \qquad b_{j,\ell} := \int_{\mathbb{R}^q} \frac{\tau_1\tau_2 \exp\left(-\sigma_X^2 \|\boldsymbol{\tau}\|^2\right)}{c_q \|\boldsymbol{\tau}\|^{q+1}} \left(\tau_1\right)^{\ell} \left(\tau_2\right)^{j-1-\ell} d\boldsymbol{\tau}, \quad \ell \textit{ odd.}$$

Diagonal terms: ℓ *even*, $q = 2$, j odd, $j = 2i + 1$, $i = 0, 1, 2, \ldots$, and $Z \sim N(0, 1)$,

$$\begin{aligned}
a_{j,\ell} &:= \int_{\mathbb{R}^q} \frac{\tau_1^2 \exp\left(-\sigma_X^2 \|\boldsymbol{\tau}\|^2\right)}{c_q \|\boldsymbol{\tau}\|^{q+1}} (\tau_1)^{\ell} (\tau_2)^{j-1-\ell} d\boldsymbol{\tau} \\
&= \frac{1}{2c_2} \lim_{r\to\infty} \int_{-r}^{r} \int_0^{2\pi} \frac{s^{j+1} \exp\left(-\sigma_X^2 s^2\right)}{s^3} \cos^{2+\ell}(\theta) \sin^{j-1-\ell}(\theta)\, s ds d\theta \\
&= \frac{1}{2c_2} \lim_{r\to\infty} \int_{-r}^{r} s^{j-1} \exp\left(-\sigma_X^2 s^2\right) ds \int_0^{2\pi} \cos^{2+\ell}(\theta) \sin^{j-1-\ell}(\theta)\, d\theta \\
&= \frac{1}{2c_2} \left\{ E\left[Z^{j-1}\right] \frac{\sqrt{\pi}}{2^{(j-1)/2}} \sigma_{\mathbf{X}}^{-j} \right\} \int_0^{2\pi} \cos^{2+\ell}(\theta) \sin^{j-1-\ell}(\theta)\, d\theta \\
&= \frac{1}{2c_2} \left\{ \left(\frac{(2i)!}{2^i i!} \right) \frac{\sqrt{\pi}}{2^i} \sigma_{\mathbf{X}}^{-1-2i} \right\} \int_0^{2\pi} \cos^{2+\ell}(\theta) \sin^{2i-\ell}(\theta)\, d\theta
\end{aligned}$$

Off-diagonal terms: ℓ *odd*, $q = 2$, j odd, $j = 2i + 1$, $i = 0, 1, 2, \ldots$,

$$\begin{aligned}
b_{j,\ell} &:= \int_{\mathbb{R}^q} \frac{\tau_1 \tau_2 \exp\left(-\sigma_X^2 \|\boldsymbol{\tau}\|^2\right)}{c_q \|\boldsymbol{\tau}\|^{q+1}} (\tau_1)^{\ell} (\tau_2)^{j-1-\ell} d\boldsymbol{\tau} \\
&= \frac{1}{2c_2} \lim_{r\to\infty} \int_{-r}^{r} \int_0^{2\pi} \frac{s^{j+1} \exp\left(-\sigma_X^2 s^2\right)}{s^3} \cos^{1+\ell}(\theta) \sin^{j-\ell}(\theta)\, s ds d\theta \\
&= \frac{1}{2c_2} \lim_{r\to\infty} \int_{-r}^{r} s^{j-1} \exp\left(-\sigma_X^2 s^2\right) ds \int_0^{2\pi} \cos^{1+\ell}(\theta) \sin^{j-\ell}(\theta)\, d\theta \\
&= \frac{1}{2c_2} \left\{ \left(\frac{(2i)!}{2^i i!} \right) \frac{\sqrt{\pi}}{2^i} \sigma_{\mathbf{X}}^{-1-2i} \right\} \int_0^{2\pi} \cos^{1+\ell}(\theta) \sin^{2i+1-\ell}(\theta)\, d\theta
\end{aligned}$$

Then $A_{j,\ell}^2$ is diagonal with typical element equal to $a_{j,\ell}^2$ or $b_{j,\ell}^2$ for ℓ even or odd, respectively, j always odd.

Now, for m, n even,

$$\int_0^{2\pi} \sin^n x \cos^m x dx = 2\pi \frac{(n-1)!!\,(m-1)!!}{(n+m)!!}$$

with the double factorial evaluated using

$$\begin{aligned}
a &= 2k \text{ even:} \qquad a!! = 2^k k! \\
a &= 2k-1 \text{ odd:} \quad a!! = \frac{(2k)!}{2^k k!},
\end{aligned}$$

so that, ℓ even, $\ell = 0, 2, \ldots, 2i$

$$\begin{aligned}\int_0^{2\pi} \cos^{2+\ell}(\theta)\sin^{2i-\ell}(\theta)\,d\theta &= 2\pi\frac{(2+\ell-1)!!\,(2i-\ell-1)!!}{(2+2i)!!}\\ &= 2\pi\frac{1}{2^{2(1+i)}}\frac{(2+\ell)!}{(1+\ell/2)!}\frac{(2i-\ell)!}{(i-\ell/2)!}\frac{1}{(1+i)!}\end{aligned}$$

and, ℓ odd, $\ell = 1, 3, \ldots, 2i-1$ $(j = 2i+1)$, $\ell := 2k-1$, so $2i-\ell = 2i-2k+1 = 2(i-k+1)-1$,

$$\begin{aligned}\int_0^{2\pi} \cos^{1+\ell}(\theta)\sin^{j-\ell}(\theta)\,d\theta &= 2\pi\frac{\ell!!\,(j-\ell-1)!!}{(j+1)!!} = 2\pi\frac{\ell!!\,(2i-\ell)!!}{(2i+2)!!}\\ &= 2\pi\frac{1}{2^{2(i+1)}}\frac{(\ell+1)!}{((\ell+1)/2)!}\frac{(2i-\ell+1)!}{(i-(\ell-1)/2)!}\frac{1}{(i+1)!}.\end{aligned}$$

Then, $c_2 = 2\pi$,

$$\begin{aligned}a_{j,\ell} &= \frac{1}{2c_2}\left\{\left(\frac{(2i)!}{2^i i!}\right)\frac{\sqrt{\pi}}{2^i}\sigma_{\mathbf{X}}^{-1-2i}\right\}2\pi\frac{1}{2^{2(1+i)}}\frac{(2+\ell)!}{(1+\ell/2)!}\frac{(2i-\ell)!}{(i-\ell/2)!}\frac{1}{(1+i)!}\\ &= \frac{\sigma_{\mathbf{X}}^{-1-2i}}{2}\frac{\sqrt{\pi}}{2^{2+4i}}\frac{(2i)!}{i!}\frac{1}{(1+i)!}\frac{(2+\ell)!}{(1+\ell/2)!}\frac{(2i-\ell)!}{(i-\ell/2)!}\end{aligned}$$

so for $\ell = 2k$ even, $j = 2i+1$ odd,

$$\begin{aligned}\sum_{\ell=0,even}^{j-1}\binom{j-1}{\ell}a_{j,\ell}^2 &= \sum_{k=0}^{i}\binom{2i}{2k}a_{2i+1,2k}^2 = \left\{\frac{\sigma_{\mathbf{X}}^{-1-2i}}{4}\frac{\sqrt{\pi}}{4}\frac{(2i)!}{i!\,(1+i)!}\right\}^2\sum_{k=0}^{i}\alpha_{2i+1,2k},\\ \alpha_{2i+1,2k} &= \frac{(2i)!}{(2k)!(2i-2k)!}\left(\frac{(2+2k)!}{(1+k)!}\frac{(2i-2k)!}{(i-k)!}\right)^2.\end{aligned}$$

Similarly,

$$\begin{aligned}b_{j,\ell} &= \frac{1}{2c_2}\left\{\left(\frac{(2i)!}{2^i i!}\right)\frac{\sqrt{\pi}}{2^i}\sigma_{\mathbf{X}}^{-1-2i}\right\}2\pi\frac{1}{2^{2(i+1)}}\frac{(\ell+1)!}{((\ell+1)/2)!}\frac{(2i-\ell+1)!}{(i-(\ell-1)/2)!}\frac{1}{(i+1)!}\\ &= \frac{\sigma_{\mathbf{X}}^{-1-2i}}{2}\frac{\sqrt{\pi}}{2^{2+4i}}\frac{(2i)!}{i!}\frac{1}{(i+1)!}\frac{(\ell+1)!}{((\ell+1)/2)!}\frac{(2i-\ell+1)!}{(i-(\ell-1)/2)!}\end{aligned}$$

and for $\ell = 2k+1$ odd,

$$\begin{aligned}\sum_{\ell=0,odd}^{j-1}\binom{j-1}{\ell}b_{j,\ell}^2 &= \sum_{k=0}^{i-1}\binom{2i}{2k+1}b_{2i+1,2k+1}^2 = \left\{\frac{\sigma_{\mathbf{X}}^{-1-2i}}{4}\frac{\sqrt{\pi}}{2^{2+4i}}\frac{(2i)!}{i!\,(1+i)!}\right\}^2\sum_{k=0}^{i-1}\beta_{2i+1,2k+1},\\ \beta_{2i+1,2k+1} &= \frac{(2i)!}{(2k+1)!(2i-2k-1)!}\left(\frac{(2(k+1))!}{(k+1)!}\frac{(2(i-k))!}{(i-k)!}\right)^2.\end{aligned}$$

Then, $j = 2i + 1$,

$$\begin{aligned}
\mathbf{A} &= \mathbf{I}_q \sum_{j=1}^{\infty} \frac{\sigma_X^{2j}}{j!} \sum_{\ell=0}^{j-1} \binom{j-1}{\ell} A_{j,\ell}^2 \\
&= \mathbf{I}_q \sum_{i=0}^{\infty} \frac{\sigma_X^{4i+2}}{(2i+1)!} \left\{ \sum_{k=0}^{i} \binom{2i}{2k} a_{2i+1,2k}^2 + \sum_{k=0}^{i-1} \binom{2i}{2k+1} b_{2i+1,2k+1}^2 \right\} \\
&= \mathbf{I}_q \sum_{i=0}^{\infty} \frac{\sigma_X^{4i+2}}{(2i+1)!} \left\{ \frac{\sigma_{\mathbf{X}}^{-1-2i}}{4} \frac{\sqrt{\pi}}{2^{2+4i}} \frac{(2i)!}{i!\,(1+i)!} \right\}^2 \left\{ \sum_{k=0}^{i} \alpha_{2i+1,2k} + \sum_{k=0}^{i-1} \beta_{2i+1,2k+1} \right\} \\
&= \mathbf{I}_q \frac{\pi}{4^2} \sum_{i=0}^{\infty} \frac{1}{2^{2+8i}} \frac{1}{(2i+1)!} \left\{ \frac{(2i)!}{i!\,(1+i)!} \right\}^2 \left\{ \sum_{k=0}^{i} \alpha_{2i+1,2k} + \sum_{k=0}^{i-1} \beta_{2i+1,2k+1} \right\},
\end{aligned}$$

where we can approximate numerically the two infinite series as $1.0287 + 5.8917 \times 10^{-3} = 1.034\,6$ so that

$$\mathbf{\Omega} = \sigma_\varepsilon^2 \Sigma_{\mathbf{YX}} \mathbf{A} \Sigma_{\mathbf{XY}} \approx 1.0346 \cdot \sigma_\varepsilon^2 \frac{\pi}{4^2} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XY}}$$

and with $\mathbf{\Upsilon} = \frac{\sqrt{\pi}}{4} \sigma_{\mathbf{X}}^{-1} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XY}}$, the asymptotic variance of $WCIV$ for $q = 2$ is

$$\mathbf{\Upsilon}^{-1} \mathbf{\Omega} \mathbf{\Upsilon}^{-1} \approx 1.0346 \cdot \sigma_\varepsilon^2 \sigma_{\mathbf{X}}^2 \left(\Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XY}} \right)^{-1}$$

which is proportional by a factor $ARE_{WCIV} = 1.0346$ to the asymptotic variance of the 2SLS estimate under homoskedasticity and iid-ness of $\mathbf{X}$,

$$\left(\Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}} \right)^{-1} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} E\left[\varepsilon^2 \mathbf{X}\mathbf{X}' \right] \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}} \left(\Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}} \right)^{-1} = \sigma_\varepsilon^2 \sigma_{\mathbf{X}}^2 \left(\Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XY}} \right)^{-1},$$

implying an even smaller efficiency loss than in the $q = 1$ case with respect to 2SLS. We conjecture in the light of our numerical experiments that the small efficiency loss of $WCIV$ shrinks as q increases.

9.1.2 Normal case with Gaussian Kernel

For $q = 1$

$$\begin{aligned}
\boldsymbol{\Upsilon} &= \sigma^2_{\mathbf{YX}} \int_{\mathbb{R}^q} \boldsymbol{\tau}^2 \exp\left(-\boldsymbol{\tau}^2 \sigma^2_{\mathbf{X}}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\boldsymbol{\tau}^2\right) d\boldsymbol{\tau} \\
&= \sigma^2_{\mathbf{YX}} \left(1+2\sigma^2_{\mathbf{X}}\right)^{-1/2} \frac{1}{\left(2\pi\left(1+2\sigma^2_{\mathbf{X}}\right)^{-1}\right)^{1/2}} \int_{\mathbb{R}^q} \boldsymbol{\tau}^2 \exp\left(-\frac{1}{2}\boldsymbol{\tau}^2\left(1+2\sigma^2_{\mathbf{X}}\right)\right) d\boldsymbol{\tau} \\
&= \sigma^2_{\mathbf{YX}} \left(1+2\sigma^2_{\mathbf{X}}\right)^{-3/2} = \rho^2_{\mathbf{YX}} \frac{\sigma^2_{\mathbf{Y}}\sigma^2_{\mathbf{X}}}{\left(1+2\sigma^2_{\mathbf{X}}\right)^{3/2}},
\end{aligned}$$

while for $Z \sim N(0,1)$,

$$\begin{aligned}
\mathbf{A} &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \boldsymbol{\tau}\boldsymbol{\lambda}' \exp\left(-\sigma^2_X \left\|\boldsymbol{\tau}\right\|^2 - \sigma^2_X \left\|\boldsymbol{\lambda}\right\|^2\right) \left[\exp\left(\sigma^2_X \boldsymbol{\tau}'\boldsymbol{\lambda}\right) - 1\right] \\
&\quad \times \exp\left(-\frac{1}{2}\left(\left\|\boldsymbol{\tau}\right\|^2 + \left\|\boldsymbol{\lambda}\right\|^2\right)\right) d\boldsymbol{\tau} d\boldsymbol{\lambda} \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \boldsymbol{\tau}\boldsymbol{\lambda} \exp\left(-\frac{1}{2}\left(1+2\sigma^2_X\right)\boldsymbol{\tau}^2 - \frac{1}{2}\left(1+2\sigma^2_X\right)\boldsymbol{\lambda}^2\right) \left[\exp\left(\sigma^2_X \boldsymbol{\tau}\boldsymbol{\lambda}\right) - 1\right] d\boldsymbol{\tau} d\boldsymbol{\lambda} \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \boldsymbol{\tau}\boldsymbol{\lambda} \exp\left(-\frac{1}{2}\left(1+2\sigma^2_X\right)\boldsymbol{\tau}^2 - \frac{1}{2}\left(1+2\sigma^2_X\right)\boldsymbol{\lambda}^2\right) \sum_{j=1}^{\infty} \frac{1}{j!} \left(\sigma^2_X \boldsymbol{\tau}\boldsymbol{\lambda}\right)^j d\boldsymbol{\tau} d\boldsymbol{\lambda} \\
&= \frac{1}{2\pi} \sum_{j=0}^{\infty} \frac{\left(\sigma^2_X\right)^{2j+1}}{(2j+1)!} \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2}\left(1+2\sigma^2_X\right)\boldsymbol{\tau}^2 - \frac{1}{2}\left(1+2\sigma^2_X\right)\boldsymbol{\lambda}^2\right) \left(\boldsymbol{\tau}\boldsymbol{\lambda}\right)^{2j+2} d\boldsymbol{\tau} d\boldsymbol{\lambda} \\
&= \left(1+2\sigma^2_X\right)^{-1} \sum_{j=0}^{\infty} \frac{\left(\sigma^2_X\right)^{2j+1}}{(2j+1)!} \\
&\quad \times \left(\frac{1}{\sqrt{2\pi\left(1+2\sigma^2_X\right)^{-1}}} \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2}\left(1+2\sigma^2_X\right)\boldsymbol{\tau}^2\right) \boldsymbol{\tau}^{2j+2} d\boldsymbol{\tau}\right)^2 \\
&= \left(1+2\sigma^2_X\right)^{-1} \sum_{j=0}^{\infty} \frac{\left(\sigma^2_X\right)^{2j+1}}{(2j+1)!} \left(1+2\sigma^2_X\right)^{-2j-2} E\left[Z^{2j+2}\right]^2,
\end{aligned}$$

which, using that $E\left[Z^{2j+2}\right]=2^{-j-1}\frac{(2j+2)!}{(j+1)!}$, is

$$\begin{aligned}
\mathbf{A} &= \left(1+2\sigma_X^2\right)^{-2}\sum_{j=0}^{\infty}\left(\frac{\sigma_X^2}{1+2\sigma_X^2}\right)^{2j+1}\frac{1}{(2j+1)!}\left[2^{-j-1}\frac{(2j+2)!}{(j+1)!}\right]^2 \\
&= \left(1+2\sigma_X^2\right)^{-2}\frac{\frac{\sigma_X^2}{1+2\sigma_X^2}}{\left(1-\left(\frac{\sigma_X^2}{1+2\sigma_X^2}\right)^2\right)^{\frac{3}{2}}} \\
&= \frac{\sigma_X^2}{(1+\sigma_X^2)^{\frac{3}{2}}\,(1+3\sigma_X^2)^{\frac{3}{2}}}.
\end{aligned}$$

Then

$$\boldsymbol{\Omega}=\sigma_\varepsilon^2\Sigma_{\mathbf{YX}}\mathbf{A}\Sigma_{\mathbf{XY}}=\sigma_\varepsilon^2\rho_{\mathbf{YX}}^2\sigma_X^2\sigma_Y^2\frac{\sigma_X^2}{(1+\sigma_X^2)^{\frac{3}{2}}\,(1+3\sigma_X^2)^{\frac{3}{2}}}$$

so that

$$\begin{aligned}
\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Omega}\boldsymbol{\Upsilon}^{-1} &= \left(\rho_{\mathbf{YX}}^2\frac{\sigma_{\mathbf{Y}}^2\sigma_{\mathbf{X}}^2}{(1+2\sigma_{\mathbf{X}}^2)^{3/2}}\right)^{-2}\sigma_\varepsilon^2\rho_{\mathbf{YX}}^2\sigma_X^2\sigma_Y^2\frac{\sigma_X^2}{(1+\sigma_X^2)^{\frac{3}{2}}\,(1+3\sigma_X^2)^{\frac{3}{2}}} \\
&= \frac{\sigma_\varepsilon^2}{\rho_{\mathbf{YX}}^2\sigma_{\mathbf{Y}}^2}\frac{(1+2\sigma_{\mathbf{X}}^2)^3}{(1+\sigma_X^2)^{\frac{3}{2}}\,(1+3\sigma_X^2)^{\frac{3}{2}}},
\end{aligned}$$

which is minimum as $\sigma_{\mathbf{X}}^2\to 0$ achieving the usual IV asymptotic variance, and in general

$$ARE_{WMD}=\frac{(1+2\sigma_{\mathbf{X}}^2)^3}{(1+\sigma_X^2)^{\frac{3}{2}}\,(1+3\sigma_X^2)^{\frac{3}{2}}}\in\left(1,\quad\frac{8}{9}\sqrt{3}\right)\approx(1,\quad 1.5396)\,,$$

which can be substantially larger than the efficiency of $WCIV$ for $q=1$, $\pi/3=1.0472$, except for $\sigma_{\mathbf{X}}^2<0.26688$. Otherwise, i.e., for larger $\sigma_{\mathbf{X}}^2$, including $\sigma_{\mathbf{X}}^2=1$, Gaussian-kernel is (possibly much) less efficient (under a Gaussian+homoskedasticity assumption) than the non-integrable kernel.

For general q,

$$\begin{aligned}
\boldsymbol{\Upsilon} &= \int_{\mathbb{R}^q}\Sigma_{\mathbf{YX}}\boldsymbol{\tau}\boldsymbol{\tau}'\Sigma_{\mathbf{XY}}\left|\varphi_{\mathbf{X}}(\boldsymbol{\tau})\right|^2\boldsymbol{\omega}(d\boldsymbol{\tau}) \\
&= \Sigma_{\mathbf{YX}}\left\{\int_{\mathbb{R}^q}\boldsymbol{\tau}\boldsymbol{\tau}'\exp\left(-\boldsymbol{\tau}'\Sigma_{\mathbf{XX}}\boldsymbol{\tau}\right)\boldsymbol{\omega}(d\boldsymbol{\tau})\right\}\Sigma_{\mathbf{XY}} \\
&= \Sigma_{\mathbf{YX}}\left\{(2\pi)^{-q/2}\int_{\mathbb{R}^q}\boldsymbol{\tau}\boldsymbol{\tau}'\exp\left(-\frac{1}{2}\left\|\boldsymbol{\tau}\right\|^2\left(1+2\sigma_{\mathbf{X}}^2\right)\right)d\boldsymbol{\tau}\right\}\Sigma_{\mathbf{XY}}
\end{aligned}$$

where for $q = 2$ the integral is diagonal with typical element equal to

$$
\begin{aligned}
& \frac{1}{2\pi}\int_{\mathbb{R}^2} \boldsymbol{\tau}_1^2 \exp\left(-\frac{1}{2}\|\boldsymbol{\tau}\|^2\left(1+2\sigma_{\mathbf{X}}^2\right)\right) d\boldsymbol{\tau} \\
= & \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} s^2 \exp\left(-\frac{1}{2}s^2\left(1+2\sigma_{\mathbf{X}}^2\right)\right) \cos^2\theta s d\theta \\
= & \frac{\pi}{2\pi}\int_0^\infty s^3 \exp\left(-\frac{1}{2}s^2\left(1+2\sigma_{\mathbf{X}}^2\right)\right) ds \\
= & \frac{1}{2}\left(2\pi\left(1+2\sigma_{\mathbf{X}}^2\right)^{-1}\right)^{1/2}\left(1+2\sigma_{\mathbf{X}}^2\right)^{-3/2}\frac{1}{2}E\left[|Z|^3\right] \\
= & 2^{-3/2}\pi^{1/2}\left(1+2\sigma_{\mathbf{X}}^2\right)^{-2}\frac{2^{3/2}}{\sqrt{\pi}} \\
= & \left(1+2\sigma_{\mathbf{X}}^2\right)^{-2},
\end{aligned}
$$

or, alternatively, this is

$$
\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \boldsymbol{\tau}_1^2 \exp\left(-\frac{1}{2}\boldsymbol{\tau}_1^2\left(1+2\sigma_{\mathbf{X}}^2\right)\right) d\boldsymbol{\tau}_1 \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\boldsymbol{\tau}_2^2\left(1+2\sigma_{\mathbf{X}}^2\right)\right) d\boldsymbol{\tau}_2 = \left(1+2\sigma_{\mathbf{X}}^2\right)^{-2}.
$$

Next, for $q = 2$,

$$
\mathbf{A} = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \boldsymbol{\tau}\boldsymbol{\lambda}' \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\|\boldsymbol{\tau}\|^2 - \frac{1}{2}\left(1+2\sigma_X^2\right)\|\boldsymbol{\lambda}\|^2\right)\left[\exp\left(\sigma_X^2\boldsymbol{\tau}'\boldsymbol{\lambda}\right)-1\right] d\boldsymbol{\tau} d\boldsymbol{\lambda}
$$

where

$$
\begin{aligned}
\boldsymbol{\tau}\boldsymbol{\lambda}'\left[\exp\left(\sigma_X^2\boldsymbol{\tau}'\boldsymbol{\lambda}\right)-1\right] &= \boldsymbol{\tau}\boldsymbol{\lambda}'\sum_{j=1}^\infty \frac{\left(\sigma_X^2\boldsymbol{\tau}'\boldsymbol{\lambda}\right)^j}{j!} \\
&= \boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\lambda}\boldsymbol{\lambda}'\sum_{j=1}^\infty \frac{\sigma_X^{2j}\left(\boldsymbol{\tau}'\boldsymbol{\lambda}\right)^{j-1}}{j!} \\
&= \boldsymbol{\tau}\boldsymbol{\tau}'\boldsymbol{\lambda}\boldsymbol{\lambda}'\sum_{i=0}^\infty \frac{\sigma_X^{2(2i+1)}\left(\boldsymbol{\tau}'\boldsymbol{\lambda}\right)^{2i}}{(2i+1)!}
\end{aligned}
$$

and for $q = 2$ and $j = 1, 2, \ldots$ (only odd j contributes, $j = 2i+1$, $i = 0, 1, \ldots$),

$$
\left(\boldsymbol{\tau}'\boldsymbol{\lambda}\right)^{j-1} = \left(\tau_1\lambda_1 + \tau_2\lambda_2\right)^{j-1} = \sum_{\ell=0}^{j-1}\binom{j-1}{\ell}\left(\tau_1\right)^\ell\left(\tau_2\right)^{j-1-\ell}\left(\lambda_1\right)^\ell\left(\lambda_2\right)^{j-1-\ell}.
$$

Then

$$\begin{aligned}
\mathbf{A} &= \sum_{j=1}^{\infty} \frac{\sigma_X^{2j}}{j!} \sum_{\ell=0}^{j-1} \binom{j-1}{\ell} A_{j,\ell}^2, \\
A_{j,\ell} &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \boldsymbol{\tau}\boldsymbol{\tau}' \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\|\boldsymbol{\tau}\|^2\right)(\tau_1)^{\ell}(\tau_2)^{j-1-\ell}\, d\tau_1 d\tau_2,
\end{aligned}$$

where we evaluate the typical elements in the 2×2 matrix $A_{j,\ell}$, diagonal and off-diagonal,

$$\begin{aligned}
&\text{diagonal:} \quad a_{j,\ell} := \frac{1}{2\pi} \int_{\mathbb{R}^2} \tau_1^2 \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\|\boldsymbol{\tau}\|^2\right)(\tau_1)^{\ell}(\tau_2)^{j-1-\ell}\, d\tau_1 d\tau_2, \quad \ell \textit{ even} \\
&\text{off-diagonal:} \quad b_{j,\ell} := \frac{1}{2\pi} \int_{\mathbb{R}^2} \tau_1\tau_2 \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\|\boldsymbol{\tau}\|^2\right)(\tau_1)^{\ell}(\tau_2)^{j-1-\ell}\, d\tau_1 d\tau_2, \quad \ell \textit{ odd.}
\end{aligned}$$

Diagonal terms: ℓ *even*, $q=2$, j odd, $j=2i+1$, $i=0,1,2,\ldots$, and $Z \sim N(0,1)$,

$$\begin{aligned}
a_{j,\ell} &:= \frac{1}{2\pi} \int_{\mathbb{R}^2} \tau_1^2 \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\|\boldsymbol{\tau}\|^2\right)(\tau_1)^{\ell}(\tau_2)^{j-1-\ell}\, d\tau_1 d\tau_2 \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \tau_1^{2+\ell} \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\tau_1^2\right) d\tau_1 \int_{\mathbb{R}} \tau_2^{j-1-\ell} \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\tau_2^2\right) d\tau_2 \\
&= \left(1+2\sigma_X^2\right)^{-1}\left(1+2\sigma_X^2\right)^{\frac{-j-1}{2}} E\left[Z^{2+\ell}\right] E\left[Z^{j-1-\ell}\right] \\
&= \left(1+2\sigma_X^2\right)^{\frac{-j-3}{2}} \left(\frac{(2+\ell)!}{2^{1+\ell/2}(1+\ell/2)!}\right)\left(\frac{(j-1-\ell)!}{2^{(j-1-\ell)/2}((j-1-\ell)/2)!}\right) \\
&= \left(1+2\sigma_X^2\right)^{-i-2} \left(\frac{(2+\ell)!}{2^{1+\ell/2}(1+\ell/2)!}\right)\left(\frac{(2i-\ell)!}{2^{i-\ell/2}(i-\ell/2)!}\right).
\end{aligned}$$

Off-diagonal terms: ℓ *odd*, $q=2$, j odd, $j=2i+1$, $i=0,1,2,\ldots$,

$$\begin{aligned}
b_{j,\ell} &:= \frac{1}{2\pi} \int_{\mathbb{R}^2} \tau_1\tau_2 \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\|\boldsymbol{\tau}\|^2\right)(\tau_1)^{\ell}(\tau_2)^{j-1-\ell}\, d\tau_1 d\tau_2 \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \tau_1^{1+\ell} \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\tau_1^2\right) d\tau_1 \int_{\mathbb{R}} \tau_2^{j-\ell} \exp\left(-\frac{1}{2}\left(1+2\sigma_X^2\right)\tau_2^2\right) d\tau_2 \\
&= \left(1+2\sigma_X^2\right)^{-1}\left(1+2\sigma_X^2\right)^{\frac{-j-1}{2}} E\left[Z^{1+\ell}\right] E\left[Z^{j-\ell}\right] \\
&= \left(1+2\sigma_X^2\right)^{\frac{-j-3}{2}} \left(\frac{(1+\ell)!}{2^{(1+\ell)/2}((1+\ell)/2)!}\right)\left(\frac{(j-\ell)!}{2^{(j-\ell)/2}((j-\ell)/2)!}\right) \\
&= \left(1+2\sigma_X^2\right)^{-i-2} \left(\frac{(1+\ell)!}{2^{(1+\ell)/2}((1+\ell)/2)!}\right)\left(\frac{(2i+1-\ell)!}{2^{i+(1-\ell)/2}(i+(1-\ell)/2)!}\right).
\end{aligned}$$

Then $A_{j,\ell}^2$ is diagonal with typical element equal to $a_{j,\ell}^2$ or $b_{j,\ell}^2$ for ℓ even or odd, respectively, j always odd.

Then, for $\ell = 2k$ even, $j = 2i+1$ odd,

$$\begin{aligned}\sum_{\ell=0,even}^{j-1} \binom{j-1}{\ell} a_{j,\ell}^2 &= \sum_{k=0}^{i} \binom{2i}{2k} a_{2i+1,2k}^2 = \left(1+2\sigma_X^2\right)^{-2i-4} \sum_{k=0}^{i} \alpha_{2i+1,2k}, \\ \alpha_{2i+1,2k} &= \frac{(2i)!}{(2k)!(2i-2k)!} \left(\frac{(2+2k)!}{2^{1+k}(1+k)!}\right)^2 \left(\frac{(2i-2k)!}{2^{i-k}(i-k)!}\right)^2.\end{aligned}$$

Similarly, for $\ell = 2k+1$ odd,

$$\begin{aligned}\sum_{\ell=0,odd}^{j-1} \binom{j-1}{\ell} b_{j,\ell}^2 &= \sum_{k=0}^{i-1} \binom{2i}{2k+1} b_{2i+1,2k+1}^2 = \left(1+2\sigma_X^2\right)^{-2i-4} \sum_{k=0}^{i-1} \beta_{2i+1,2k+1}, \\ \beta_{2i+1,2k+1} &= \frac{(2i)!}{(2k+1)!(2i-2k-1)!} \left(\frac{(2+2k)!}{2^{k+1}(k+1)!}\right)^2 \left(\frac{(2i-2k)!}{2^{i-k}(i-k)!}\right)^2.\end{aligned}$$

Then,

$$\begin{aligned}\mathbf{A} &= \sum_{j=1}^{\infty} \frac{\sigma_X^{2j}}{j!} \sum_{\ell=0}^{j-1} \binom{j-1}{\ell} A_{j,\ell}^2 \\ &= \mathbf{I}_q \sum_{i=0}^{\infty} \frac{\sigma_X^{4i+2}}{(2i+1)!} \left\{ \sum_{k=0}^{i} \binom{2i}{2k} a_{2i+1,2k}^2 + \sum_{k=0}^{i-1} \binom{2i}{2k+1} b_{2i+1,2k+1}^2 \right\} \\ &= \mathbf{I}_q \sum_{i=0}^{\infty} \frac{\sigma_X^{4i+2} \left(1+2\sigma_X^2\right)^{-2i-4}}{(2i+1)!} \left\{ \sum_{k=0}^{i} \alpha_{2i+1,2k} + \sum_{k=0}^{i-1} \beta_{2i+1,2k+1} \right\} \\ &= \mathbf{I}_q \frac{\sigma_X^2}{\left(1+2\sigma_X^2\right)^4} \sum_{i=0}^{\infty} \frac{1}{(2i+1)!} \frac{\left(\sigma_X^2\right)^{2i}}{\left(1+2\sigma_X^2\right)^{2i}} \left\{ \sum_{k=0}^{i} \alpha_{2i+1,2k} + \sum_{k=0}^{i-1} \beta_{2i+1,2k+1} \right\} \\ &= \mathbf{I}_q \frac{\sigma_X^2}{\left(1+2\sigma_X^2\right)^4} \gamma_q \left(\sigma_{\mathbf{X}}^2\right)\end{aligned}$$

where we can approximate numerically the two infinite series in $ARE_{WMD} = \gamma_2\left(\sigma_{\mathbf{X}}^2\right) \in (1,\ 1.778)$ for given values of σ_X^2, e.g. for $\sigma_{\mathbf{X}}^2 \in \{1/2, 1, 2\}$ we have $\gamma_2\left(\sigma_X^2\right) = \{1.137\,8, 1.265\,6, 1.417\,2\}$

Then, for $q = 2$,

$$\mathbf{\Omega} = \sigma_\varepsilon^2 \Sigma_{\mathbf{YX}} \mathbf{A} \Sigma_{\mathbf{XY}} = \gamma_2\left(\sigma_X^2\right) \frac{\sigma_X^2}{\left(1+2\sigma_X^2\right)^4} \sigma_\varepsilon^2 \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XY}}$$

so that with $\mathbf{\Upsilon} = \left(1+2\sigma_{\mathbf{X}}^2\right)^{-2} \Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XY}}$, the asymptotic variance of $WCIV$ for $q = 2$ is

$$\mathbf{\Upsilon}^{-1}\mathbf{\Omega}\mathbf{\Upsilon}^{-1} = \gamma_2\left(\sigma_X^2\right) \cdot \sigma_\varepsilon^2 \sigma_{\mathbf{X}}^2 \left(\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XY}}\right)^{-1} = ARE_{WMD} \cdot \sigma_\varepsilon^2 \sigma_{\mathbf{X}}^2 \left(\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XY}}\right)^{-1}$$

so WMD can be very inefficient compared to 2SLS estimation if $\sigma_{\mathbf{X}}^2$ is not very small, where the inefficiency can be larger the larger is q.

9.1.3 Exponential case with non-integrable kernel

1. We model the joint distribution of $\mathbf{Y}$ and $\mathbf{X}$ for $q = 1$ with the Bivariate Exponential distribution of Marshall and Olkin (1967), with joint moment generating function given by, $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$,

$$\psi(\mu, \tau) = \frac{(\lambda + \mu + \tau)(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12}) + \mu\tau\lambda_{12}}{(\lambda + \mu + \tau)(\lambda_1 + \lambda_{12} + \mu)(\lambda_2 + \lambda_{12} + \tau)}$$

so that the characteristic function is

$$\varphi_{\mathbf{Y},\mathbf{X}}(\mu, \tau) = \frac{(\lambda - i\mu - i\tau)(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12}) - \mu\tau\lambda_{12}}{(\lambda - i\mu - i\tau)(\lambda_1 + \lambda_{12} - i\mu)(\lambda_2 + \lambda_{12} - i\tau)}$$

and if $\lambda_{12} = 0$, Y and X become independent with exponential marginals

$$\varphi_X(\tau) = \varphi(0, \tau) = \frac{(\lambda - i\tau)(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12})}{(\lambda - i\tau)(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12} - i\tau)} = \frac{\lambda_2}{\lambda_2 - i\tau} \sim Exp(\lambda_2),$$

and in general the marginals are exponential

$$\varphi_X(\tau) = \varphi(0, \tau) = \frac{\lambda_2 + \lambda_{12}}{\lambda_2 + \lambda_{12} - i\tau} \sim Exp(\lambda_2 + \lambda_{12}) \sim Exp\left(m_2^{-1}\right).$$

Since Y is centered, we focus on, $m_1 = \frac{1}{\lambda_1 + \lambda_{12}}$,

$$\varphi_{\mathbf{Z}}(\mu, \tau) = \varphi_{Y - m_1, X}(\mu, \tau) = \varphi(\mu, \tau)\, e^{-i\mu m_1}.$$

Then,

$$\frac{\partial}{\partial \boldsymbol{\mu}} \varphi_{\mathbf{Z}}(\mu, \tau) = \frac{\partial}{\partial \mu}\left(\frac{(\lambda - i\mu - i\tau)(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12}) - \mu\tau\lambda_{12}}{(\lambda - i\mu - i\tau)(\lambda_1 + \lambda_{12} - i\mu)(\lambda_2 + \lambda_{12} - i\tau)}\right) - i m_1 e^{-i\mu m_1} \varphi(\mu, \tau),$$

and

$$\begin{aligned} \left.\frac{\partial}{\partial \boldsymbol{\mu}} \varphi_{\mathbf{Z}}(\boldsymbol{\mu}, \boldsymbol{\tau})\right|_{\mu = \mathbf{0}} &= \frac{(\lambda_1 + \lambda_{12})\left(i\lambda^2\lambda_2 - i\tau^2\lambda_2 + i\lambda^2\lambda_{12} + 2\lambda\tau\lambda_2 + \lambda\tau\lambda_{12}\right)}{(\lambda_1 + \lambda_{12})^2 (\lambda - i\tau)^2 (\lambda_2 - i\tau + \lambda_{12})} - i m_1 \frac{(\lambda_2 + \lambda_{12})}{(\lambda_2 + \lambda_{12} - i\tau)} \\ &= \frac{i\left(\lambda^2\lambda_2 + \lambda^2\lambda_{12} - \tau^2\lambda_2\right) + (2\lambda_2 + \lambda_{12})\lambda\tau}{(\lambda_1 + \lambda_{12})(\lambda - i\tau)^2 (\lambda_2 + \lambda_{12} - i\tau)} - i m_1 \frac{(\lambda_2 + \lambda_{12})}{(\lambda_2 + \lambda_{12} - i\tau)} \end{aligned}$$

so this is

$$
\begin{aligned}
&= \frac{i\left(\left(\lambda^2\lambda_2+\lambda^2\lambda_{12}-\tau^2\lambda_2\right)-m_1\left(\lambda_2+\lambda_{12}\right)\left(\lambda_1+\lambda_{12}\right)\left(\lambda-i\tau\right)^2\right)+\left(2\lambda_2+\lambda_{12}\right)\lambda\tau}{\left(\lambda_1+\lambda_{12}\right)\left(\lambda-i\tau\right)^2\left(\lambda_2+\lambda_{12}-i\tau\right)} \\
&= \frac{\lambda_{12}\tau\left\{i\tau-\lambda\right\}}{\left(\lambda_1+\lambda_{12}\right)\left(\lambda-i\tau\right)^2\left(\lambda_2+\lambda_{12}-i\tau\right)} \\
&= -\tau\sigma_{YX}\frac{\lambda}{\lambda-i\tau}\varphi_{\mathbf{X}}\left(\tau\right)
\end{aligned}
$$

and

$$
\left|\frac{\partial}{\partial\boldsymbol{\mu}}\varphi_{\mathbf{Z}}\left(\boldsymbol{\mu},\boldsymbol{\tau}\right)\Big|_{\mu=\mathbf{0}}\right|^2=\frac{\lambda_{12}^2\tau^2}{\left(\lambda_1+\lambda_{12}\right)^2\left(\lambda^2+\tau^2\right)\left(\left(\lambda_2+\lambda_{12}\right)^2+\tau^2\right)}.
$$

Then, for the non-integrable kernel we have,

$$
\begin{aligned}
\boldsymbol{\Upsilon} &= \int_{\mathbb{R}}\left\{\frac{\partial}{\partial\boldsymbol{\mu}}\varphi_{\mathbf{Z}}\left(\boldsymbol{\mu},\boldsymbol{\tau}\right)\Big|_{\mu=\mathbf{0}}\right\}\left\{\frac{\partial}{\partial\boldsymbol{\mu}}\varphi_{\mathbf{Z}}\left(\mu,-\boldsymbol{\tau}\right)\Big|_{\mu=\mathbf{0}}\right\}\boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \\
&= \int_{\mathbb{R}}\left|\frac{\partial}{\partial\boldsymbol{\mu}}\varphi_{\mathbf{Z}}\left(\boldsymbol{\mu},\boldsymbol{\tau}\right)\Big|_{\mu=\mathbf{0}}\right|^2\frac{1}{\pi\tau^2}d\tau \\
&= \frac{\lambda_{12}^2m_1^2}{\pi}\int_{\mathbb{R}}\frac{1}{\left(\lambda^2+\tau^2\right)\left(m_2^{-2}+\tau^2\right)}d\tau \\
&= \rho_{YX}^2\sigma_Y^2\lambda^2\frac{1}{\pi}\int_{\mathbb{R}}\frac{1}{\left(\lambda^2+\tau^2\right)\left(m_2^{-2}+\tau^2\right)}d\tau \\
&= \rho_{YX}^2\sigma_X^2\sigma_Y^2\frac{1}{\pi}\int_{\mathbb{R}}\frac{1}{\left(1+(\tau/\lambda)^2\right)\left(1+\left(\tau m_2\right)^2\right)}d\tau, \quad m_2=\sigma_X \\
&= \rho_{YX}^2\sigma_Y^2\sigma_X\frac{1}{\pi}\int_{\mathbb{R}}\frac{1}{\left(1+(x/\lambda m_2)^2\right)\left(1+x^2\right)}dx, \\
&= \rho_{YX}^2\sigma_Y^2\sigma_X\frac{\lambda m_2}{\lambda m_2+1}=\rho_{YX}^2\sigma_Y^2\sigma_X\frac{\lambda}{\lambda+m_2^{-1}} \\
&= \rho_{YX}^2\sigma_Y^2\sigma_X\left\{\frac{\lambda_1+\lambda_2+\lambda_{12}}{\lambda_1+2\lambda_2+2\lambda_{12}}\right\}=\rho_{YX}^2\sigma_Y^2\sigma_X\left\{1+\frac{m_2^{-1}}{\lambda}\right\}^{-1}
\end{aligned}
$$

where

$$
\sigma_{YX}=\frac{\lambda_{12}}{\lambda\left(\lambda_1+\lambda_{12}\right)\left(\lambda_2+\lambda_{12}\right)}=\left\{\frac{\lambda_{12}}{\lambda}\right\}m_1m_2=\left\{\rho_{YX}\right\}\sigma_X\sigma_Y.
$$

We can set

$$
\boldsymbol{\Upsilon}=\rho_{YX}^2\sigma_Y^2\sigma_X\cdot c_{\boldsymbol{\Upsilon}}\left(\rho,m_1,m_2\right)
$$

$$
c_{\boldsymbol{\Upsilon}}\left(\rho,m_1,m_2\right):=\left\{1+\frac{m_2^{-1}}{\lambda}\right\}^{-1}=\left\{1+\frac{m_2^{-1}}{\lambda_{12}}\rho\right\}^{-1}=\left\{1+\frac{m_2^{-1}}{\left(m_1^{-1}+m_2^{-1}\right)}\left(\rho+1\right)\right\}^{-1}
$$

Then, for comparison purposes we fix

$$\rho = \frac{\lambda_{12}}{\lambda} \quad \text{and} \quad \sigma_Y^2 = \sigma_1^2 = (\lambda_1 + \lambda_{12})^{-2} = m_1^2$$

and keep free $m_2 = (\lambda_2 + \lambda_{12})^{-1} = \sigma_2 = \sigma_X$.

For the variance term we obtain

$$\begin{aligned}
E\left[\frac{\partial}{\partial \boldsymbol{\beta}} h(\boldsymbol{\beta}_0, \boldsymbol{\tau})\right] &= -E\left[(\mathbf{Y}_t - \boldsymbol{\mu}_Y)\exp(i\langle \boldsymbol{\tau}, \mathbf{X}_t\rangle)\right] = -\frac{1}{i}\frac{\partial}{\partial \boldsymbol{\mu}}\varphi_{\mathbf{Z}}(\boldsymbol{\mu}, \boldsymbol{\tau})\Big|_{\boldsymbol{\lambda}=\mathbf{0}} \\
&= -\frac{\lambda_{12}\tau\{i\tau - \lambda\}}{(\lambda_1 + \lambda_{12})(\lambda - i\tau)^2(\lambda_2 + \lambda_{12} - i\tau)} \\
&= \frac{\lambda_{12}\tau}{(\lambda_2 + \lambda_{12})(\lambda_1 + \lambda_{12})(\lambda - i\tau)}\frac{\lambda_{12} + \lambda_2}{(\lambda_2 + \lambda_{12} - i\tau)} \\
&= \tau\frac{\lambda_{12}}{\lambda(\lambda_2 + \lambda_{12})(\lambda_1 + \lambda_{12})}\frac{\lambda}{(\lambda - i\tau)}\varphi_{\mathbf{X}}(\tau) \\
&= \tau\sigma_{YX}\frac{\lambda}{(\lambda - i\tau)}\varphi_{\mathbf{X}}(\tau).
\end{aligned}$$

Then, under homoskedasticity,

$$\begin{aligned}
\frac{\boldsymbol{\Omega}}{\sigma_\varepsilon^2\sigma_{YX}^2} &= \int_{\mathbb{R}^2}\tau\mu\frac{\lambda}{(\lambda - i\tau)}\frac{\lambda}{(\lambda + i\mu)}\varphi_{\mathbf{X}}(\tau)\varphi_{\mathbf{X}}(-\mu)\left[\varphi_{\mathbf{X}}(\mu - \tau) - \varphi_{\mathbf{X}}(\mu)\varphi_{\mathbf{X}}(-\tau)\right]\omega(d\tau)\,\omega(d\mu) \\
&= \int_{\mathbb{R}^2}\tau\mu\frac{\lambda}{(\lambda - i\tau)}\frac{\lambda}{(\lambda + i\mu)}\left[\varphi_{\mathbf{X}}(\tau)\varphi_{\mathbf{X}}(-\mu)\varphi_{\mathbf{X}}(\mu - \tau) - |\varphi_{\mathbf{X}}(\tau)|^2|\varphi_{\mathbf{X}}(\mu)|^2\right]\omega(d\tau)\,\omega(d\mu)
\end{aligned}$$

where for the univariate Exponential case,

$$\varphi_{\mathbf{X}}(\tau) = \frac{\lambda_{12} + \lambda_2}{\lambda_{12} + \lambda_2 - i\tau}, \quad |\varphi_{\mathbf{X}}(\tau)|^2 = \frac{(\lambda_{12} + \lambda_2)^2}{(\lambda_{12} + \lambda_2)^2 + \tau^2},$$

so that $\mathbf{\Omega}/\left(\sigma_{\varepsilon}^{2}\sigma_{YX}^{2}\right)$ is, $m=1/\lambda$,

$$
\begin{aligned}
&\int_{\mathbb{R}^2}\frac{\tau\mu}{\pi^2\tau^2\mu^2}\frac{\lambda}{(\lambda-i\tau)}\frac{\lambda}{(\lambda+i\mu)}\left[\varphi_{\mathbf{X}}(-\mu)\varphi_{\mathbf{X}}(\tau)\varphi_{\mathbf{X}}(\mu-\tau)-|\varphi_{\mathbf{X}}(\mu)|^2|\varphi_{\mathbf{X}}(\tau)|^2\right]d\tau d\mu\\
=\;&\frac{1}{\pi^2}\int_{\mathbb{R}^2}\frac{1}{\tau\mu}\frac{\lambda}{(\lambda-i\tau)}\frac{\lambda}{(\lambda+i\mu)}\\
&\times\left[\frac{\lambda_{12}+\lambda_2}{\lambda_{12}+\lambda_2+i\mu}\frac{\lambda_{12}+\lambda_2}{\lambda_{12}+\lambda_2-i\tau}\frac{\lambda_{12}+\lambda_2}{\lambda_{12}+\lambda_2-i(\mu-\tau)}-\frac{(\lambda_{12}+\lambda_2)^2}{(\lambda_{12}+\lambda_2)^2+\mu^2}\frac{(\lambda_{12}+\lambda_2)^2}{(\lambda_{12}+\lambda_2)^2+\tau^2}\right]d\tau d\mu\\
=\;&\frac{1}{\pi^2}\int_{\mathbb{R}^2}\frac{1}{\tau\mu}\frac{1}{1-i\tau m}\frac{1}{1+i\mu m}\\
&\times\left[\frac{1}{1+i\mu m_2}\frac{1}{1-i\tau m_2}\frac{1}{1-i(\mu-\tau)m_2}-\frac{1}{1+(\mu m_2)^2}\frac{1}{1+(\tau m_2)^2}\right]d\tau d\mu\\
=\;&\frac{1}{\pi^2}\int_{\mathbb{R}^2}\frac{1}{\tau\mu}\frac{1+i\tau m}{1+(\tau m)^2}\frac{1-i\mu m}{1+(\mu m)^2}\\
&\times\left[\frac{1-i\mu m_2}{1+(\mu m_2)^2}\frac{1+i\tau m_2}{1+(\tau m_2)^2}\frac{1+i(\mu-\tau)m_2}{1+((\mu-\tau)m_2)^2}-\frac{1}{1+(\mu m_2)^2}\frac{1}{1+(\tau m_2)^2}\right]d\tau d\mu\\
=\;&\frac{1}{\pi^2}\int_{\mathbb{R}^2}\frac{1}{\tau\mu}\frac{1+i\tau m}{1+(\tau m)^2}\frac{1-i\mu m}{1+(\mu m)^2}\frac{(1-i\mu m_2)(1+i\tau m_2)(1+i(\mu-\tau)m_2)-\left(1+((\mu-\tau)m_2)^2\right)}{\left(1+(\mu m_2)^2\right)\left(1+(\tau m_2)^2\right)\left(1+((\mu-\tau)m_2)^2\right)}d\tau d\mu\\
=\;&\frac{m_2^2}{\pi^2}\int_{\mathbb{R}^2}\frac{1}{1+(\tau m)^2}\frac{1}{1+(\mu m)^2}\frac{((m^2-2mm_2)\tau\mu+m\tau^2m_2+m\mu^2m_2+1)}{\left(1+(\mu m_2)^2\right)\left(1+(\tau m_2)^2\right)\left(1+((\mu-\tau)m_2)^2\right)}d\tau d\mu,
\end{aligned}
$$

which can be evaluated numerically at different combinations of (m,m_1,m_2) or equivalently of $(\lambda_{12},\lambda_1,\lambda_2)$, see Table 9.

Then we can evaluate the ARE of $WCIV$ with respect to IV as

$$\mathbf{\Upsilon}^{-1}\mathbf{\Omega}\mathbf{\Upsilon}^{-1}=ARE_{WCIV}\frac{\sigma_{\varepsilon}^{2}}{\rho_{YX}^{2}\sigma_{Y}^{2}},$$

using the numerical approximation of $\mathbf{\Omega}/\left(\sigma_{\varepsilon}^{2}\sigma_{YX}^{2}\right)=\mathbf{\Omega}/\left\{\sigma_{\varepsilon}^{2}\rho_{YX}^{2}\sigma_{Y}^{2}\sigma_{X}^{2}\right\}$, where the next table shows that the $WCIV$ can be more efficient than the IV estimate.

$\{\lambda_{12},\lambda_1,\lambda_2\}$	$\{\frac{1}{2},\frac{1}{2},\frac{1}{2}\}$	$\{\frac{1}{2},\frac{1}{2},\frac{1}{4}\}$	$\{\frac{1}{2},\frac{1}{2},\frac{1}{8}\}$	$\{\frac{1}{2},\frac{1}{2},1\}$	$\{\frac{1}{2},\frac{1}{2},2\}$
$\lambda = m^{-1}$	1.5	$\frac{5}{4}$	$\frac{9}{8}$	2	3
$\rho_{YX}^2 = \lambda_{12}/\lambda$	1/3	$\frac{2}{5}$	$\frac{4}{9}$	$\frac{1}{4}$	$\frac{1}{6}$
σ_Y^2	1	1	1	1	1
$\sigma_X = m_2 = (\lambda_{12}+\lambda_{12})^{-1}$	1	$\frac{4}{3}$	$\frac{8}{5}$	$\frac{2}{3}$	$\frac{2}{5}$
$\sigma_X^2 = m_2^2 = (\lambda_{12}+\lambda_{12})^{-2}$	1	$\frac{16}{9}$	$\frac{64}{25}$	$\frac{4}{9}$	$\frac{4}{25}$
$\boldsymbol{\Upsilon}/\{\rho_{YX}^2\sigma_Y^2\sigma_X\} = \frac{\lambda_1+\lambda_2+\lambda_{12}}{\lambda_1+2\lambda_2+2\lambda_{12}}$	3/5	$\frac{5}{8}$	$\frac{9}{14}$	$\frac{4}{7}$	$\frac{6}{11}$
$\boldsymbol{\Omega}/\{\sigma_\varepsilon^2\rho_{YX}^2\sigma_Y^2\sigma_X^2\} = \frac{\boldsymbol{\Omega}}{\sigma_\varepsilon^2\sigma_{YX}^2}$	0.32571	0.33781	0.34581	0.31070	0.29606
ARE_{WCIV}	0.90475	0.86479	0.83677	0.95152	0.99509

Table 9: $WCIV$ ARE for (Y,X) exponential under homoskedasticity, $q=1$. Non-integrable kernel.

9.1.4 Exponential case with Gaussian Kernel

Now we have

$$\begin{aligned}
\boldsymbol{\Upsilon} &= \int_{\mathbb{R}} \alpha(\tau)\,\tau^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\tau^2\right) d\tau \\
&= \rho_{YX}^2\sigma_X^2\sigma_Y^2 \int_{\mathbb{R}} \frac{\tau^2}{\left(1+(\tau/\lambda)^2\right)\left(1+(\tau m_2)^2\right)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\tau^2\right) d\tau \\
&= \rho_{YX}^2\sigma_X^2\sigma_Y^2\sigma_X^{-2} \int_{\mathbb{R}} \frac{x^2}{\left(1+(x/\lambda m_2)^2\right)\left(1+x^2\right)} \frac{1}{\sqrt{2\pi m_2^2}} \exp\left(-\frac{1}{2}\left(\frac{x}{m_2}\right)^2\right) dx \\
&= \rho_{YX}^2\sigma_X^{-1}\sigma_Y^2 \int_{\mathbb{R}} \frac{x^2}{\left(1+(x/\lambda m_2)^2\right)\left(1+x^2\right)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x}{m_2}\right)^2\right) dx
\end{aligned}$$

while for the variance term we have

$$\frac{\boldsymbol{\Omega}}{\sigma_\varepsilon^2\sigma_{YX}^2} = \int_{\mathbb{R}^2} \tau\mu \frac{\lambda}{(\lambda-i\tau)}\frac{\lambda}{(\lambda+i\mu)} \left[\varphi_{\mathbf{X}}(\tau)\,\varphi_{\mathbf{X}}(-\mu)\,\varphi_{\mathbf{X}}(\mu-\tau) - |\varphi_{\mathbf{X}}(\tau)|^2|\varphi_{\mathbf{X}}(\mu)|^2\right] \omega(d\tau)\,\omega(d\mu)$$

where for the univariate Exponential case,

$$\varphi_{\mathbf{X}}(\tau) = \frac{\lambda_{12}+\lambda_2}{\lambda_{12}+\lambda_2-i\tau}, \quad |\varphi_{\mathbf{X}}(\tau)|^2 = \frac{(\lambda_{12}+\lambda_2)^2}{(\lambda_{12}+\lambda_2)^2+\tau^2}.$$

$\{\lambda_{12},\lambda_1,\lambda_2\}$	$\left\{\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\}$	$\left\{\frac{1}{2},\frac{1}{2},\frac{1}{4}\right\}$	$\left\{\frac{1}{2},\frac{1}{2},\frac{1}{8}\right\}$	$\left\{\frac{1}{2},\frac{1}{2},1\right\}$	$\left\{\frac{1}{2},\frac{1}{2},2\right\}$
$\mathbf{\Upsilon}/\left\{\rho^2_{YX}\sigma^2_Y\sigma_X\right\}$	0.212 48	0.185 87	0.167 46	0.236 62	0.229 98
$\mathbf{\Omega}/\left\{\sigma^2_\varepsilon\rho^2_{YX}\sigma^2_Y\sigma^2_X\right\}$	4.4084×10^{-2}	3.2403×10^{-2}	2.5317×10^{-2}	5.6012×10^{-2}	5.2998×10^{-2}
ARE_{WMD}	0.97644	0.93792	0.90280	1.0004	1.0020
$ARE_{WMD/WCIV}$	1.0792	1.0846	1.0789	1.051 4	1.0069

Table 10: WMD ARE for (Y,X) exponential under homoskedasticity, $q=1$. Gaussian kernel.

Then we have that $\mathbf{\Omega}/\left(\sigma^2_\varepsilon\sigma^2_{YX}\right)=\mathbf{\Omega}/\left\{\sigma^2_\varepsilon\rho^2_{YX}\sigma^2_Y\sigma^2_X\right\}$ is, $m=1/\lambda$,

$$\begin{aligned}
&\int_{\mathbb{R}^2}\frac{\tau\mu}{2\pi}\frac{\lambda}{(\lambda-i\tau)}\frac{\lambda}{(\lambda+i\mu)}\left[\begin{array}{c}\varphi_{\mathbf{X}}(-\mu)\,\varphi_{\mathbf{X}}(\tau)\,\varphi_{\mathbf{X}}(\mu-\tau)\\ -|\varphi_{\mathbf{X}}(\mu)|^2|\varphi_{\mathbf{X}}(\tau)|^2\end{array}\right]\exp\left(-\frac{1}{2}\left(\tau^2+\mu^2\right)\right)d\tau d\mu\\
=\;&\int_{\mathbb{R}^2}\frac{\tau\mu}{2\pi}\frac{\lambda}{(\lambda-i\tau)}\frac{\lambda}{(\lambda+i\mu)}\exp\left(-\frac{1}{2}\left(\tau^2+\mu^2\right)\right)\\
&\times\left[\frac{\lambda_{12}+\lambda_2}{\lambda_{12}+\lambda_2+i\mu}\frac{\lambda_{12}+\lambda_2}{\lambda_{12}+\lambda_2-i\tau}\frac{\lambda_{12}+\lambda_2}{\lambda_{12}+\lambda_2-i(\mu-\tau)}-\frac{(\lambda_{12}+\lambda_2)^2}{(\lambda_{12}+\lambda_2)^2+\mu^2}\frac{(\lambda_{12}+\lambda_2)^2}{(\lambda_{12}+\lambda_2)^2+\tau^2}\right]d\tau d\mu\\
=\;&\int_{\mathbb{R}^2}\frac{\tau\mu}{2\pi}\frac{1+i\tau m}{1+(\tau m)^2}\frac{1-i\mu m}{1+(\mu m)^2}\\
&\times\left[\frac{(1-i\mu m_2)(1+i\tau m_2)(1+i(\mu-\tau)m_2)-\left(1+((\mu-\tau)m_2)^2\right)}{\left(1+(\mu m_2)^2\right)\left(1+(\tau m_2)^2\right)\left(1+((\mu-\tau)m_2)^2\right)}\right]\exp\left(-\frac{1}{2}\left(\tau^2+\mu^2\right)\right)d\tau d\mu\\
=\;&m_2^2\int_{\mathbb{R}^2}\frac{(\tau\mu)^2}{2\pi}\frac{1}{1+(\tau m)^2}\frac{1}{1+(\mu m)^2}\\
&\times\frac{(m^2\tau\mu+m\tau^2m_2-2m\tau\mu m_2+m\mu^2m_2+1)}{\left(1+(\mu m_2)^2\right)\left(1+(\tau m_2)^2\right)\left(1+((\mu-\tau)m_2)^2\right)}\exp\left(-\frac{1}{2}\left(\tau^2+\mu^2\right)\right)d\tau d\mu\\
=\;&m_2^2\int_{\mathbb{R}^2}\frac{(\tau\mu)^2}{2\pi}\frac{1}{1+(\tau m)^2}\frac{1}{1+(\mu m)^2}\\
&\times\frac{((m^2-2mm_2)\tau\mu+mm_2(\tau^2+\mu^2)+1)}{\left(1+(\mu m_2)^2\right)\left(1+(\tau m_2)^2\right)\left(1+((\mu-\tau)m_2)^2\right)}\exp\left(-\frac{1}{2}\left(\tau^2+\mu^2\right)\right)d\tau d\mu,
\end{aligned}$$

which can be evaluated numerically for given values of $\{m,m_1,m_2\}$ or $\{\lambda_{12},\lambda_1,\lambda_2\}$, see Table 10.

Then, the asymptotic variance of WMD with respect to IV is obtained as

$$\mathbf{\Upsilon}^{-1}\mathbf{\Omega}\mathbf{\Upsilon}^{-1}=ARE_{WMD}\frac{\sigma^2_\varepsilon}{\rho^2_{YX}\sigma^2_Y},$$

and we can similarly obtain the ARE with respect to $WCIV$. We find that WMD can be sometimes more efficient than IV (when σ_X is large), but is uniformly less efficient than $WCIV$, though the differences are smaller than for Gaussian data.

		WNIV	WNIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$\phi = 0$					$q = 3$			
	Med	0.0019	0.0021	0.0165	0.0201	0.0587	0.0823	0.1037
	DecR	1.5514	1.5380	2.1156	2.0141	1.0070	1.3113	1.8193
	Rej	0.0585	0.0586	0.0643	0.0653	0.0325	0.0644	0.1030
					$q = 10$			
	Med	0.0232	0.0297	0.1845	0.2275	0.0800	0.1176	0.1901
	DecR	2.0877	1.9185	6.7898	2.7965	1.2626	2.0299	2.7117
	Rej	0.0700	0.0716	0.1027	0.1117	0.0397	0.0946	0.1297
					$q = 15$			
	Med	0.0284	0.0403	0.3657	0.4755	0.0890	0.1471	0.2417
	DecR	2.4397	2.0094	8.6042	1.1919	1.4153	2.3481	2.9748
	Rej	0.0728	0.0751	0.1400	0.2336	0.0453	0.1134	0.1376
$\phi = 0.5$					$q = 3$			
	Med	−0.0135	−0.0132	−0.0109	−0.0072	0.0712	0.0929	0.1233
	DecR	1.6605	1.6436	2.1331	2.0507	1.0885	1.4635	1.9495
	Rej	0.0468	0.0468	0.0554	0.0562	0.0308	0.0728	0.1111
					$q = 10$			
	Med	0.0040	0.0102	0.1709	0.2154	0.0812	0.1404	0.2036
	DecR	2.1980	1.9844	7.2131	2.7872	1.2888	2.1450	2.7175
	Rej	0.0583	0.0595	0.0996	0.1094	0.0407	0.1024	0.1346
					$q = 15$			
	Med	0.0098	0.0223	0.3251	0.4676	0.0838	0.1558	0.2540
	DecR	2.5774	2.1109	9.2929	1.1594	1.4222	2.3938	2.9931
	Rej	0.0648	0.0676	0.1300	0.2188	0.0476	0.1243	0.1436

Table 11: Linear IV model $M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t}$, $Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t$. 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. Sample size is 500. The elements of $\mathbf{X}_t$ are pairwise independent.

9.2 More Monte Carlo Simulations

In this subsection, we report simulation results regarding M_1 considered in Section 6 to illustrate the poor finite-sample properties of WMD and WMDF in the case that the elements of $\mathbf{X}_t$ are pairwise independent. The sample size now is 500. From Table 11, it is observed that the finite-sample properties of WMD and WMDF deteriorate severely when q increases, while WCIV and WCIVF maintain excellent finite-sample properties in all the cases.

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. (1982). Consistent model specification tests. *Journal of Econometrics*, 20:105–134.

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.

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. High Dimensional Problems in Econometrics.

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.

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.

Lyons, R. (2013). Distance covariance in metric spaces. *The Annals of Probability*, 41(5):3284 – 3305.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

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.

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.

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.

Stinchcombe, M. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 14:295 – 325.

Székely, G. J. and Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of Classification*, 22:151–183.

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. (2024). Generalized spectral tests for multivariate martingale difference hypotheses. *Journal of Business & Economic Statistics*, forthcoming.

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.