Consistent Parameter Estimation for Conditional Moment Restrictions

Shih-Hsun Hsu
Department of Economics
National Taiwan University

Chung-Ming Kuan
Institute of Economics
Academia Sinica

† Author for correspondence: Chung-Ming Kuan, Institute of Economics, Academia Sinica, Taipei 115, Taiwan; ckuan@econ.sinica.edu.tw.

†† We thank John Chao, Ying-Ying Lee, Halbert White, and Yongmiao Hong for very useful comments and suggestions. The research support from the National Science Council of the Republic of China (NSC95-2415-H-001-034) for C.-M. Kuan is gratefully acknowledged.

Abstract

In estimating conditional moment restrictions, a well known difficulty is that the estimator based on a set of implied unconditional moments may lose its consistency when the parameters of interest are not globally identified. In this paper, we consider a continuum of unconditional moments that are equivalent to the postulated conditional moments and can identify the parameters of interest. We propose to project these unconditional moments along the exponential Fourier series and construct an objective function based on the resulting Fourier coefficients. A novel feature of our method is that the full continuum of unconditional moments is incorporated into each Fourier coefficient. We show that, when the number of Fourier coefficients in the objective function grows at a proper rate, the proposed estimator is consistent and asymptotically normally distributed. Our simulations confirm that the proposed estimator compares favorably with that of Domínguez and Lobato (2004, *Econometrica*) in terms of bias, standard error and mean squared error. For models with exogenous regressors, the proposed estimator may also outperform the nonlinear least squares estimator when there are multiple local minima.

JEL classification: C12, C22

Keywords: conditional moment restrictions, Fourier coefficients, generically comprehensive revealing function, global identifiability, GMM

1 Introduction

Many economic and econometric models can be characterized in terms of conditional moment restrictions. Consistent estimation of the parameters in such restrictions is thus a crucial step in empirical studies. A typical approach, hereafter the unconditional moment approach, is to find a finite set of unconditional moment restrictions implied by the original, conditional restrictions and apply a suitable estimation method, such as the generalized method of moment (GMM) of Hansen (1982) and Hansen and Singleton (1982), or the empirical likelihood method of Qin and Lawless (1994) and Kitamura (1997). The instrumental-variable estimation method for regression models is a leading example. Although there are nonparametric methods that deal with the conditional moments directly, e.g., Ai and Chen (2003) and Kitamura, Tripathi, and Ahn (2004), the unconditional moment approach remains popular in practice.

A critical assumption in the unconditional moment approach is that the parameters in the conditional restrictions can be globally identified by the implied, unconditional restrictions. With this assumption, estimator consistency is not really an issue and can be easily established under suitable regularity conditions. Therefore, much research interest focuses on estimator efficiency, e.g., Chamberlain (1987), Newey (1990, 1993), Carrasco and Florens (2000), and Donald, Imbens, and Newey (2003). The assumption of global identifiability was challenged by Domínguez and Lobato (2004). They showed that this assumption easily fails to hold in nonlinear models so that estimator consistency can not be guaranteed. The identification problem arises because the unconditional moments are usually chosen arbitrarily and hence may not help to identify the parameters of interest. We provide a simple example as follows to demonstrate this problem.

Assume that the random variable Y satisfies the simple nonlinear model $\mathbb{E}[Y|X] = \exp(X\theta_o)$, where θ_o is an unknown parameter of interest and X follows a normal distribution with a zero mean and an nonzero variance. The corresponding conditional moment restriction is then given by $\mathbb{E}[Y - \exp(X\theta_o)|X] = 0$. Based on this conditional restriction, we consider three valid, implied unconditional moment restrictions to illustrate the identification problem, they are, respectively, (i) $\mathbb{E}[Y - \exp(X\theta)] = 0$; (ii) $\mathbb{E}[(Y - \exp(X\theta))X] = 0$; and (iii) $\mathbb{E}[(Y - \exp(X\theta))X^2] = 0$. Some calculations show that restrictions (i) and (iii) can not help to identify θ_o since $-\theta_o$ also satisfies these two restrictions. Instead, moment restriction (ii) gives the unique solution θ_o and thus is helpful to identify θ_o. Domínguez and Lobato (2004) also provided some examples and demonstrated that the identification problem may happen even when the unconditional moments are resulted from the so-called optimal instruments.

To circumvent the potential problems of identifiability and inconsistency, Domínguez and Lobato (2004) proposed implementing the unconditional moment approach by finding a continuum of unconditional moment restrictions that are equivalent to the original, conditional restrictions. In particular, they constructed unconditional moments by employing the "instruments" generated from an indicator function. There are some disadvantages of their method, however. First, the indicator function takes only the values one and zero and hence may not well present the information in the conditioning variables. Second, their estimation method does not utilize the full continuum of moment restrictions. This may result in further efficiency loss, as argued by Carrasco and Florens (2000).

This paper proposes a new class of consistent estimators for conditional moment restrictions. Our estimator differs from that of Domínguez and Lobato (2004) in the following respects. First, we consider a continuum of equivalent unconditional moment restrictions with "instruments" resulted from a "generically comprehensively revealing" (GCR) function (Stinchcombe and White, 1998), such that they can globally identify the parameters of interest. Second, instead of working on these moments directly, we project them along the exponential Fourier series and construct an objective function based on the resulting Fourier coefficients. This leads to a consistent estimator. A novel feature of our method is that it in effect utilizes all possible information in the conditioning variables because all unconditional moments have been incorporated into *each* Fourier coefficient. Compared with Carrasco and Florens (2000), our method is practically simpler because we do not require preliminary estimation of the eigenvalues and eigen-functions of a covariance operator.

It is shown in this paper that the proposed estimator is consistent and asymptotically normally distributed when the number of Fourier coefficients in the objective function grows at a proper rate. We also specialize on the "instruments" obtained from the exponential function, a special case in the class of GCR functions. With these instruments, the resulting unconditional moments and Fourier coefficients have analytic forms which greatly facilitate estimation in practice. Our simulations confirm that, under various settings, the proposed estimator compares favorably with that of Domínguez and Lobato (2004) in terms of bias, standard error and mean squared error. It is also found that, even for models with exogenous regressors, the proposed estimator may deliver smaller bias and mean squared error than does the nonlinear least squares estimator when there are multiple local minima.

This paper is organized as follows. We introduce the proposed estimator in Section 2 and establish its asymptotic properties in Section 3. The simulation results are reported in Section 4. Section 5 concludes this paper. All proofs are deferred to the Appendix.

2 The Proposed Estimator

We are interested in estimating $\boldsymbol{\theta}_o$, the $q \times 1$ vector of unknown parameters, in the following conditional moment restriction:

$$\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o)|\boldsymbol{X}] = \mathbf{0}, \quad \text{with probability one (w.p.1)}, \tag{1}$$

where $\boldsymbol{h}$ is a $p \times 1$ vector of functions, $\boldsymbol{Y}$ is a $r \times 1$ vector of data variables, and $\boldsymbol{X}$ is an $m \times 1$ vector of conditioning variables. Without loss of generality, we shall work on the case that $\boldsymbol{X}$ is bounded with probability one; see e.g., Bierens (1994, Theorem 3.2.1).

It is well known that (1) is equivalent to the unconditional moment restriction:

$$\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) f(\boldsymbol{X})] = \mathbf{0}, \tag{2}$$

for all measurable functions f, where each $f(\boldsymbol{X})$ may be interpreted as an "instrument" that helps to identify $\boldsymbol{\theta}_o$. In practice, it is infeasible to consider all possible functions. Thus, one typically forms an estimating function by subjectively choosing certain instruments, such as the square and cross product of the elements in $\boldsymbol{X}$. This would not be a problem in a linear model if the resulting unconditional moments can exactly identify θ_o. Yet, when $\boldsymbol{h}$ is nonlinear in $\boldsymbol{\theta}_o$, Domínguez and Lobato (2004) showed that $\boldsymbol{\theta}_o$ is not necessarily identified when unconditional moments are determined arbitrarily, and its identifiability may depend on the marginal distributions of the conditioning variables $\boldsymbol{X}$. This concern is practically relevant because models with nonlinear restrictions are quite common in econometric applications; see e.g., Hansen and Singleton (1982) and Hansen and West (2002).[1]

One way to ensure parameter identifiability is to employ a class of instruments that span a space of functions of $\boldsymbol{X}$ (Bierens, 1982, 1990; Stinchcombe and White, 1998). Domínguez and Lobato (2004) did this by setting the instruments as $\mathbf{1}(\boldsymbol{X} \leq \boldsymbol{\tau}) = \prod_{j=1}^{m} \mathbf{1}(X_j \leq$

[1] Hansen and West (2002) studied the papers published in 7 top economics journals in 1990 and 2000 and found that, among 35 articles that employed the GMM technique, 14 of them deal with models with nonlinear restrictions.

τ_j), where $\mathbf{1}(B)$ is the indicator function of the event B. This leads to a continuum of unconditional moments indexed by $\boldsymbol{\tau}$ that are equivalent to (1):

$$\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o)\mathbf{1}(\boldsymbol{X}\le\boldsymbol{\tau})]=\mathbf{0},\quad \boldsymbol{\tau}\in\mathbb{R}^m. \tag{3}$$

Then, $\boldsymbol{\theta}_o$ can be globally identified by an L_2-norm of these moments, i.e.,

$$\boldsymbol{\theta}_o=\underset{\boldsymbol{\theta}\,\in\,\Theta}{\operatorname{argmin}}\int_{\mathbb{R}^m}\left|\mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\mathbf{1}(\boldsymbol{X}\le\boldsymbol{\tau})\big]\right|^2\,\mathrm{d}P(\boldsymbol{\tau}), \tag{4}$$

with $P(\boldsymbol{\tau})$ a distribution function of $\boldsymbol{\tau}$. Here, a natural choice of $P(\boldsymbol{\tau})$ is $P_{\boldsymbol{X}}(\boldsymbol{\tau})$, the distribution function of $\boldsymbol{X}$. The L_2-norm in (4) is thus an expectation with respect to $P_{\boldsymbol{X}}(\boldsymbol{\tau})$ and can be well approximated by the sample average. This leads Dimínguez and Lobato (2004) to propose the following estimator:

$$\widehat{\boldsymbol{\theta}}_{\mathrm{DL}}(T)=\underset{\boldsymbol{\theta}\,\in\,\Theta}{\operatorname{argmin}}\frac{1}{T}\sum_{k=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta})\mathbf{1}(\boldsymbol{x}_t\le\boldsymbol{\tau}_k)\right)^2, \tag{5}$$

where $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ are the sample observations of $\boldsymbol{Y}$ and $\boldsymbol{X}$, respectively, and $\boldsymbol{\tau}_k=\boldsymbol{x}_k$, $k=1,\ldots,T$. This is precisely a GMM estimator based on T unconditional moments induced by the indicator function. By the analogy between the L_2-norm in (4) and the objective function in (5), $\widehat{\boldsymbol{\theta}}_{\mathrm{DL}}(T)$ is consistent for $\boldsymbol{\theta}_o$ under regularity conditions.

2.1 A Class of Consistent Estimators

The indicator function is not the only choice for the desired instruments; Stinchcombe and White (1998) showed that any GCR function will also do. In particular, for a real analytic function G that is not a polynomial,[2] $G(A(\boldsymbol{X},\boldsymbol{\tau}))$ can serve as an instrument in (2), where A is the affine transformation such that $A(\boldsymbol{X},\boldsymbol{\tau})=\tau_0+\sum_{j=1}^{m}X_j\tau_j$. For example, G may be the exponential function (Bierens, 1982, 1990) or the logistic function (White, 1989).

A striking property of the instruments resulted from a GCR function is that (2) holds for the instruments with the index $\boldsymbol{\tau}$ in an arbitrarily chosen index set in $\mathbb{R}^{m+1}$; see Stinchcombe and White (1998, p. 304). As such, the unconditional moment restrictions induced by a GCR function are

$$\mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o)G\big(A(\boldsymbol{X},\boldsymbol{\tau})\big)\big]=\mathbf{0},\ \text{ for almost all }\boldsymbol{\tau}\in\mathcal{T}\subset\mathbb{R}^{m+1}, \tag{6}$$

[2] A function G is said to be analytic if it locally equals its Taylor expansion at every point of its domain.

where $\mathcal{T}$ may be a small subset with a nonempty interior. Note that the indicator function is not GCR; hence (3) must hold for all $\boldsymbol{\tau}$ in $\mathbb{R}^m$. Similar to (4), $\boldsymbol{\theta}_o$ now can be globally identified by the L_2-norm of (6):

$$\boldsymbol{\theta}_o = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \int_{\mathcal{T}} \big|\mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})G\big(A(\boldsymbol{X},\boldsymbol{\tau})\big)\big]\big|^2 \,\mathrm{d}P(\boldsymbol{\tau}). \tag{7}$$

In contrast with Domínguez and Lobato (2004), there is no natural choice of $P(\boldsymbol{\tau})$. It is therefore not easy to find a proper sample counterpart of the L_2-norm in (7). Although an objective function for estimating $\boldsymbol{\theta}_o$ can be constructed using randomized $\boldsymbol{\tau}$, the resulting estimate is arbitrary and may not be preferred.

In this paper, we take a different approach to deriving a class of consistent estimators for $\boldsymbol{\theta}_o$. This approach is based on a condition equivalent to the L_2-norm in (7). We project the unconditional moments in (6) along the exponential Fourier series and obtain

$$\mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})G\big(A(\boldsymbol{X},\boldsymbol{\tau})\big)\big] = \frac{1}{(2\pi)^{m+1}} \sum_{\boldsymbol{k}\in\mathcal{S}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}) \exp{(i\boldsymbol{k}'\boldsymbol{\tau})},$$

where $\mathcal{S} := \{\boldsymbol{k} = [k_0, k_1, \ldots, k_m]' \in \mathbb{Z}^{m+1}\}$ with $k_i = 0, \pm1, \pm2, \cdots, \pm\infty$, and $C_{G,\boldsymbol{k}}(\boldsymbol{\theta})$ is a Fourier coefficient:

$$\begin{aligned} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}) &= \int_{\mathcal{T}} \mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})G\big(A(\boldsymbol{X},\boldsymbol{\tau})\big)\big] \exp{(-i\boldsymbol{k}'\boldsymbol{\tau})} \,\mathrm{d}\boldsymbol{\tau} \\ &= \mathbb{E}\left[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}) \int_{\mathcal{T}} G\big(A(\boldsymbol{X},\boldsymbol{\tau})\big) \exp{(-i\boldsymbol{k}'\boldsymbol{\tau})} \,\mathrm{d}\boldsymbol{\tau}\right], \quad \boldsymbol{k} \in \mathcal{S}. \end{aligned}$$

It can be seen that each $C_{G,\boldsymbol{k}}(\boldsymbol{\theta})$ incorporates the continuum of the original instruments $G(A(\boldsymbol{X},\boldsymbol{\tau}))$ into a new instrument:

$$\varphi_{G,\boldsymbol{k}}(\boldsymbol{X}) = \int_{\mathcal{T}} G\big(A(\boldsymbol{X},\boldsymbol{\tau})\big) \exp{(-i\boldsymbol{k}'\boldsymbol{\tau})} \,\mathrm{d}\boldsymbol{\tau}, \tag{8}$$

in which the index parameter $\boldsymbol{\tau}$ has been integrated out. Apart from a scaling factor, Parseval's Theorem implies that the L_2-norm in (7) is equivalent to

$$\sum_{\boldsymbol{k}\in\mathcal{S}} \big|C_{G,\boldsymbol{k}}(\boldsymbol{\theta})\big|^2 = \sum_{\boldsymbol{k}\in\mathcal{S}} \big|\mathbb{E}\left[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(\boldsymbol{X})\right]\big|^2.$$

It follows that $\boldsymbol{\theta}_o$ can be identified as

$$\boldsymbol{\theta}_o = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \sum_{\boldsymbol{k}\in\mathcal{S}} \big|\mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(\boldsymbol{X})\big]\big|^2, \tag{9}$$

where the right-hand side depends only on a countable collection of unconditional moments that do not involve $\boldsymbol{\tau}$.

By replacing $\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(\boldsymbol{X})]$ in (9) with its sample counterpart, an objective function for estimating $\boldsymbol{\theta}_o$ is readily obtained. It is well known that $C_{G,\boldsymbol{k}}(\boldsymbol{\theta}) \to 0$ as $\|\boldsymbol{k}\|$ tends to infinity by Bessel's inequality. This suggests that the new instruments $\varphi_{G,\boldsymbol{k}}(\boldsymbol{X})$, and hence $\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(\boldsymbol{X})]$, contain little information for identifying $\boldsymbol{\theta}_o$ when $\|\boldsymbol{k}\|$ is large.[3] As such, we may omit "remote" Fourier coefficients and compute an estimator of $\boldsymbol{\theta}_o$ as

$$\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}}) = \underset{\boldsymbol{\theta}\,\in\,\Theta}{\operatorname{argmin}} \sum_{\boldsymbol{k}\in\mathcal{S}(\mathcal{K}_T)} \left| \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(\boldsymbol{x}_t) \right|^2, \tag{10}$$

where $\mathcal{K}_T$ grows with T but at a slower rate and $\mathcal{S}(\mathcal{K}_T)$ is a subset of $\mathcal{S}$ with $k_i = 0, \pm 1, \ldots, \pm\mathcal{K}_T$. The proposed estimator (10) depends on the function G, and it is a GMM estimator based on $(2\mathcal{K}_T + 1)^{m+1}$ unconditional moments with the identity weighting matrix.[4]

Note that the Domínguez-Lobato estimator (5) does not utilize all possible information in estimation because it relies only on a finite number of unconditional moments. By contrast, the proposed estimator (10) is free from this problem because each $\varphi_{G,\boldsymbol{k}}$ has included the full continuum of the instruments required for identifying $\boldsymbol{\theta}_o$. Our estimator is also computationally simpler than that of Carrasco and Florens (2000), which requires preliminary estimation of a covariance operator and its eigenvalues and eigen-functions. Also, a regularization parameter must be determined in practice so as to ensure the invertibility of the estimated covariance operator. Therefore, the estimator of Carrasco and Florens (2000) is not easy to implement and may be arbitrary in practice.

2.2 A Specific Estimator

To compute the proposed estimator, we must specify a G function. Following Bierens (1982, 1990), we set G as the exponential function. This choice has some advantages relative to the indicator function. First, the indicator function takes only the values one and zero, whereas the exponential function is more flexible and hence may better presents the

[3]The derived moment conditions for identifying the parameter thus are "weak" in the Stock and Wright (2000) sense.

[4]Note that this weighting matrix is not "optimal" in GMM sense, the resulted estimator is thus not most efficient. See also the Domínguez-Lobato estimator (5).

information in the conditioning variables. That is, the exponential function may generate better instruments for identifying $\boldsymbol{\theta}_o$. Second, the exponential function is smooth and hence is convenient in an optimization program. Further, $\exp(A(\boldsymbol{X},\boldsymbol{\tau}))$ with $\boldsymbol{\tau} \in \mathbb{R}^{m+1}$ and $\exp(\boldsymbol{X}'\boldsymbol{\tau})$ with $\boldsymbol{\tau} \in \mathbb{R}^m$ only differ by a constant and hence play the same role in function approximation (Stinchcombe and White, 1998). By employing $\exp(\boldsymbol{X}'\boldsymbol{\tau})$ as a desired instrument, we are able to reduce the dimension of integration in (7) by one, i.e., $\mathcal{T} \subset \mathbb{R}^m$, and the summation in (9) is over $\mathcal{S} = \{\boldsymbol{k} = [k_1, \ldots, k_m]' \in \mathbb{Z}^m\}$.

More importantly, choosing $\exp(\boldsymbol{X}'\boldsymbol{\tau})$ results in an analytic form for the instrument $\varphi_{\exp,\boldsymbol{k}}$ which facilitates estimation in practice. In particular, setting $\mathcal{T} = [-\pi, \pi]^m$, the new instruments that integrate out $\boldsymbol{\tau}$ are

$$\begin{aligned}\varphi_{\exp,\boldsymbol{k}}(\boldsymbol{X}) &= \int_{\mathcal{T}} \exp(\boldsymbol{X}'\boldsymbol{\tau}) \exp\left(-i\boldsymbol{k}'\boldsymbol{\tau}\right) \mathrm{d}\boldsymbol{\tau} \\ &= \varphi_{\exp,k_1}(X_1) \cdots \varphi_{\exp,k_m}(X_m), \quad \boldsymbol{k} \in \mathcal{S},\end{aligned}$$

where

$$\begin{aligned}\varphi_{\exp,k_j}(X_j) &= \int_{-\pi}^{\pi} \exp(X_j\tau_j) \exp(-ik_j\tau_j) \,\mathrm{d}\tau_j \\ &= \frac{(-1)^{k_j} \cdot 2\sinh\left(\pi X_j\right)}{(X_j - ik_j)}, \qquad j = 1, \ldots, m,\end{aligned}$$

and $\sinh(w) = (\exp(w) - \exp(-w))/2$. Based on $\varphi_{\exp,\boldsymbol{k}}(\boldsymbol{X})$, $\boldsymbol{\theta}_o$ can be identified as in (9). The proposed estimator now reads

$$\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}}) = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \sum_{\boldsymbol{k} \in \mathcal{S}(\mathcal{K}_T)} \left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t, \boldsymbol{\theta}) \varphi_{\exp,\boldsymbol{k}}(\boldsymbol{x}_t) \right|^2, \tag{11}$$

where $\boldsymbol{k}$ is $m \times 1$.

3 Asymptotic Properties

We now establish the asymptotic properties of the proposed estimator $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})$. To ease our illustration and proof, we first consider the case that $\boldsymbol{X}$ is univariate (i.e., $m = 1$), denoted as X. The asymptotic properties for multivariate $\boldsymbol{X}$ are given in Section 3.3.

3.1 Consistency

We impose the following conditions.

[A1] The observed data $(\boldsymbol{y}_t', x_t)'$, $t = 1, \ldots, T$, are independent realizations of $(\boldsymbol{Y}', X)'$.

[A2] For each $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{h}(\cdot, \boldsymbol{\theta})$ is measurable, and for each $\boldsymbol{y} \in \mathbb{R}^r$, $\boldsymbol{h}(\boldsymbol{y}, \cdot)$ is continuous on Θ, where Θ is a compact subset in $\mathbb{R}^q$. Also, $\boldsymbol{\theta}_o$ in Θ is the unique solution to $\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta})|\boldsymbol{X}] = \mathbf{0}$.

[A3] $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \Theta} |\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta})|^2] < \infty$.

[A4] G is real analytic but not a polynomial such that, w.p.1, $\sup_{\boldsymbol{\tau} \in \mathcal{T}} |G(A(X, \boldsymbol{\tau}))| < \infty$, $\sup_{\boldsymbol{\tau} \in \mathcal{T}} |G_i(A(X, \boldsymbol{\tau}))| < \infty$, and $\sup_{\boldsymbol{\tau} \in \mathcal{T}} |G_{ij}(A(X, \boldsymbol{\tau}))| < \infty$, where $G_i(A(X, \boldsymbol{\tau})) = \partial G(A(X, \boldsymbol{\tau}))/\partial \tau_i$ and $G_{ij}(A(X, \boldsymbol{\tau})) = \partial^2 G(A(X, \boldsymbol{\tau}))/(\partial \tau_i \partial \tau_j)$, for $i, j = \{0, 1\}$.

These conditions are convenient and quite standard in the GMM literature. They may be relaxed at the expense of more technicality. For example, it is possible to extend [A1] to allow for weakly dependent and heterogeneously distributed data; see, e.g., Gallant and White (1988) and Chen and White (1996). Note that in [A2], $\boldsymbol{\theta}_o$ is assumed to be the unique solution to the original conditional restrictions; we do *not* require $\boldsymbol{\theta}_o$ to be the unique solution to some implied, unconditional moment restrictions. As in Stinchcombe and White (1998), [A4] requires G to be real analytic but not a polynomial. [A4] also imposes additional restrictions on G, yet it still permits quite general G functions.

Setting $\mathcal{T} = [-\pi, \pi]^2$, the instruments resulted from G are

$$\varphi_{G,\boldsymbol{k}}(X) = \int_{[-\pi,\pi]^2} G\big(A(\boldsymbol{X}, \boldsymbol{\tau})\big) \exp\left(-i\boldsymbol{k}'\boldsymbol{\tau}\right) \mathrm{d}\boldsymbol{\tau}. \tag{12}$$

Here, $\boldsymbol{k} = (k_0, k_1)'$. Define $c(k_i) = |k_i|$ for $k_i \neq 0$ and $c(k_i) = 1$ for $k_i = 0$, $i = 0, 1$. The result below provides a bound on $\varphi_{G,\boldsymbol{k}}(X)$.

Lemma 3.1 *Given* [A4], $|\varphi_{G,\boldsymbol{k}}(X)| \leq \Delta/[c(k_0)c(k_1)]$ *w.p.1, where* Δ *is a real number.*

Define the sample counterpart of $C_{G,\boldsymbol{k}}(\boldsymbol{\theta})$ as

$$\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t, \boldsymbol{\theta}) \varphi_{G,\boldsymbol{k}}(x_t).$$

With Lemma 3.1, we are able to characterize the approximating capability of $\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})$.

Lemma 3.2 *Given* [A1]–[A4]*, if* $\mathcal{K}_T = o(T^{1/2})$*, then*

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T} |\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}) - C_{G,\boldsymbol{k}}(\boldsymbol{\theta})|^2 \xrightarrow{\mathbb{P}} 0,$$

uniformly in $\boldsymbol{\theta} \in \Theta$*, where* $\xrightarrow{\mathbb{P}}$ *stands for convergence in probability.*

Lemma 3.2 implies

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T} \left|\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})\right|^2 \xrightarrow{\mathbb{P}} \sum_{k_0,k_1=-\infty}^{\infty} \left|C_{G,\boldsymbol{k}}(\boldsymbol{\theta})\right|^2, \tag{13}$$

uniformly for all $\boldsymbol{\theta}$ in Θ. As $\boldsymbol{\theta}_o$ is the unique minimizer of the right-hand side of (13), the consistency result below follows from Theorem 2.1 of Newey and McFadden (1994).

Theorem 3.3 *Given* [A1]–[A4]*, if* $\mathcal{K}_T = o(T^{1/2})$*, then* $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_T) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_o$ *as* $T \to \infty$*.*

For the estimator $\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}})$ in (11), note that $\exp(X\tau)$ satisfies [A4] with τ a scalar. It is easy to deduce that Lemma 3.1 holds with $|\varphi_{\exp,k}(X)| \le \Delta/k$. In analogy with Lemma 3.2, we also have

$$\sum_{k=-\mathcal{K}_T}^{\mathcal{K}_T} |\boldsymbol{m}_{\exp,k,T}(\boldsymbol{\theta}) - C_{\exp,k}(\boldsymbol{\theta})|^2 \xrightarrow{\mathbb{P}} 0, \tag{14}$$

when $\mathcal{K}_{\mathrm{T}} = o(T)$. The result below then follows from (14) and is analogous to Theorem 3.3.

Theorem 3.4 *Given* [A1]–[A3]*, if* $\mathcal{K}_T = o(T)$*, then* $\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_T) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_o$ *as* $T \to \infty$*.*

3.2 Asymptotic Normality

We shall use the following notations. Given a complex number (function) f, let $\bar{f}$ denote its complex conjugate and $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ denote its real and imaginary parts, respectively. For a vector of complex functions $\boldsymbol{f}$, its complex conjugate, real part and imaginary part are defined elementwise. Recall also that the Fourier coefficient $C_{G,\boldsymbol{k}}(\boldsymbol{\theta})$ can be expressed as

$$\mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(\boldsymbol{X})\big] = \int_{[-\pi,\pi]^2} \mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})G\big(A(X,\boldsymbol{\tau})\big)\big] \exp(-i\boldsymbol{k}'\boldsymbol{\tau})\,\mathrm{d}\boldsymbol{\tau},$$

which is the integral of the product of two functions in $\boldsymbol{\tau}$, i.e., $\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})G(A(X,\cdot))]$ and $\exp(-i\boldsymbol{k}'\cdot)$. To establish asymptotic normality, we work on $\mathbb{E}[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})G(A(X,\cdot))]$ and its sample counterpart directly. This requires some results in the function space, as given below.

Consider functions in the Hilbert space $L_2[-\pi,\pi]$. The inner product of two $p\times 1$ vectors of functions $\boldsymbol{f}$ and $\boldsymbol{g}$ is $\langle \boldsymbol{f},\boldsymbol{g}\rangle = \int_{-\pi}^{\pi} \boldsymbol{f}(\tau)'\bar{\boldsymbol{g}}(\tau)\,\mathrm{d}\tau$, and the norm induced by the inner product is $\langle \boldsymbol{f},\boldsymbol{f}\rangle^{1/2}$. A random element $\boldsymbol{U}$ has mean $\mathbb{E}(\boldsymbol{U})$ if $\mathbb{E}[\langle \boldsymbol{U},\boldsymbol{g}\rangle] = \langle \mathbb{E}(\boldsymbol{U}),\boldsymbol{g}\rangle$ for any $\boldsymbol{g}$ in $L_2[-\pi,\pi]$. The covariance operator $\mathbb{K}$ associated with $\boldsymbol{U}$ is, for any $\boldsymbol{g}$ in $L_2[-\pi,\pi]$,

$$\mathbb{K}\boldsymbol{g} = \mathbb{E}\big[\langle \boldsymbol{U}-\mathbb{E}(\boldsymbol{U}),\, \boldsymbol{g}\rangle\big(\boldsymbol{U}-\mathbb{E}(\boldsymbol{U})\big)\big],$$

such that

$$\begin{aligned}(\mathbb{K}\boldsymbol{g})(\tau) &= \mathbb{E}\big[\langle \boldsymbol{U}-\mathbb{E}(\boldsymbol{U}),\, \boldsymbol{g}\rangle\big(\boldsymbol{U}(\tau)-\mathbb{E}(\boldsymbol{U}(\tau))\big)\big]\\ &= \left(\sum_{i=1}^{p}\int_{-\pi}^{\pi}\kappa_{ji}(\tau,s)g_i(s)\,\mathrm{d}s\right)_{j=1,\dots,p},\end{aligned}$$

with the kernel $\kappa_{ji}(\tau,s)=\mathbb{E}[(U_j(\tau)-EU_j(\tau))(U_i(s)-EU_i(s))]$. $\boldsymbol{U}$ is Gaussian if for any $\boldsymbol{g}$ in $L_2[-\pi,\pi]$, $\langle \boldsymbol{U},\boldsymbol{g}\rangle$ has a normal distribution on $\mathbb{R}$ with mean $\langle \mathbb{E}(\boldsymbol{U}),\boldsymbol{g}\rangle$ and variance $\langle \mathbb{K}\boldsymbol{g},\boldsymbol{g}\rangle$. Analogous results also hold in $L_2([-\pi,\pi]^m)$. For more discussions on random elements in Hilbert space; see, e.g., Chen and White (1998) and Carrasco and Florens (2000).

In view of (10), $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})$ must satisfy the first order condition:

$$\begin{aligned}\mathbf{0} &= \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \nabla_{\boldsymbol{\theta}}\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})'\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})'\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})\\ &= \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} 2\,\mathrm{Re}\left(\nabla_{\boldsymbol{\theta}}\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})'\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})\right),\end{aligned}$$

where $\nabla_{\boldsymbol{\theta}}\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})$ is a $p\times q$ matrix with $\nabla_{\theta_i}\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta})$ its i-th column. A mean-value expansion of $\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})\big)$ about $\boldsymbol{\theta}_o$ gives

$$\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})\big) = \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) + \nabla_{\boldsymbol{\theta}}\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}\big(\boldsymbol{\theta}_T^{\dagger}\big)\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})-\boldsymbol{\theta}_o\big),$$

where $\boldsymbol{\theta}_T^{\dagger}$ is between $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})$ and $\boldsymbol{\theta}_o$, and its value may be different for each row in the matrix $\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{G,\boldsymbol{k},T}\big(\boldsymbol{\theta}_T^{\dagger}\big)$. Thus,

$$\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \mathrm{Re}\,\Big(\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{G,\boldsymbol{k},T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})\big)' \\ \big[\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) + \nabla_{\boldsymbol{\theta}} \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}\big(\boldsymbol{\theta}_T^{\dagger}\big)\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}}) - \boldsymbol{\theta}_o\big)\big]\Big) = \boldsymbol{0}. \tag{15}$$

To derive the limiting distribution of normalized $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})$, we impose the following conditions.

[A5] $\boldsymbol{\theta}_o$ is in the interior of Θ.

[A6] For each $\boldsymbol{y}$, $\boldsymbol{h}(\boldsymbol{y},\cdot)$ is continuously differentiable in a neighborhood N of $\boldsymbol{\theta}_o$ such that $\mathbb{E}\big[\sup_{\boldsymbol{\theta}\in N} \|\nabla_{\boldsymbol{\theta}} \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\|^2\big] < \infty$, where $\|\cdot\|$ is a matrix norm.

[A7] The $q \times q$ matrix $\mathcal{M}_q$, with the (i,j)-th element

$$\Big\langle \mathbb{E}\left[\nabla_{\theta_i} \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G(A(X,\cdot))\right],\ \mathbb{E}\Big[\nabla_{\theta_j} \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G(A(X,\cdot))\Big]\Big\rangle,$$

is symmetric and positive definite.

[A8] $T^{-1/2}\sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta}_o) G(A(x_t,\cdot)) \xrightarrow{D} \mathbb{Z}$, where $\xrightarrow{D}$ denotes convergence in distribution, and $\mathbb{Z}$ is a p-dimensional Gaussian random element that has mean zero and the covariance operator $\mathbb{K}$ with

$$(\mathbb{K}\boldsymbol{g})(\tau) = \mathbb{E}\big[\langle \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G(A(X,\cdot)),\, \boldsymbol{g}\rangle \big(\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G(A(X,\tau))\big)\big],$$

for any p-dimensional function $\boldsymbol{g}$.

Here, [A5] is needed for mean-value expansion; [A6] is a standard "smoothness" condition in nonlinear models. By [A7], $\mathcal{M}_q$ is invertible so that the normalized estimator has a unique representation, as given in (16) below. We directly assume functional convergence in [A8] for convenience; this condition is the same as Assumption 11 in Carrasco and Florens (2000). One may, of course, impose more primitive conditions on $\boldsymbol{h}$, G and the data, so as to ensure such convergence; see e.g., Chen and White (1998).

To study the behavior of the normalized estimator via (15), we give two limiting results for the terms on the right-hand side of (15).

Lemma 3.5 *Given* [A1]–[A6], *if* $\mathcal{K}_T = o(T^{1/4})$, *then*

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T} \operatorname{Re}\Big(\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{G,\boldsymbol{k},T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T)\big)' \nabla_{\boldsymbol{\theta}} \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_T^{\dagger})\Big) \xrightarrow{\mathbb{P}} \sum_{k_0,k_1=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \nabla_{\boldsymbol{\theta}} \overline{C}_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o).$$

The limit in Lemma 3.5 is precisely the matrix $\mathcal{M}_q$ defined in [A7], because its (i,j)-th element is

$$\sum_{k_0,k_1=-\infty}^{\infty} \nabla_{\theta_i} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \nabla_{\theta_j} \overline{C}_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o) = \Big\langle \mathbb{E}\big[\nabla_{\theta_i} \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G\big(A(X,\cdot)\big)\big],\, \mathbb{E}\big[\nabla_{\theta_j} \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G\big(A(X,\cdot)\big)\big] \Big\rangle,$$

by the Multiplication theorem (see, e.g., Stuart, 1961).

Lemma 3.6 *Given* [A1]–[A6], *if* $\mathcal{K}_T = o(T^{1/4})$, *then*

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T} \operatorname{Re}\Big(\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{G,\boldsymbol{k},T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T)\big)' \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o)\Big) = \sum_{k_0,k_1=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1).$$

With Lemma 3.5 and Lemma 3.6, we can express (15) as

$$\sqrt{T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}}) - \boldsymbol{\theta}_o\big) = -\mathcal{M}_q^{-1} \left[\sum_{k_0,k_1=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \right] + o_{\mathbb{P}}(1). \tag{16}$$

The functional convergence condition [A8] then ensures that the term in the square bracket on the right-hand side of (16) also has a limiting normal distribution, which in turn leads to the asymptotic normality of $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})$.

Theorem 3.7 *Given* [A1]–[A8], *if* $\mathcal{K}_T = o(T^{1/4})$, *then*

$$\sqrt{T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T) - \boldsymbol{\theta}_o\big) \xrightarrow{D} \mathcal{N}(0,\mathcal{V}),$$

where $\mathcal{V} = \mathcal{M}_q^{-1} \boldsymbol{\Omega}_q \mathcal{M}_q^{-1}$ *and* $\boldsymbol{\Omega}_q$ *is a* $q \times q$ *matrix with the* (i,j)*-th element:*

$$\Big\langle \mathbb{E}\left[\nabla_{\theta_i} \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G(A(X,\cdot))\right],\ \mathbb{K}\,\mathbb{E}\Big[\nabla_{\theta_j} \boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G(A(X,\cdot))\Big] \Big\rangle.$$

For the estimator $\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}})$ with $G(A(X,\boldsymbol{\tau})) = \exp(X\tau)$, it can be verified that the results analogous to Lemma 3.5 and Lemma 3.6 hold when $\mathcal{K}_{\mathrm{T}}$ is $o(T^{1/2})$. In particular,

$$\sum_{k=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \mathrm{Re}\left(\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{\exp,k,T}\big(\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}})\big)' \nabla_{\boldsymbol{\theta}} \overline{\boldsymbol{m}}_{exp,k,T}(\boldsymbol{\theta}_T^{\dagger})\right) \xrightarrow{\mathbb{P}} \sum_{k=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{\exp,k}(\boldsymbol{\theta}_o)' \nabla_{\boldsymbol{\theta}} \overline{C}_{\exp,k}(\boldsymbol{\theta}_o), \tag{17}$$

which is the matrix $\mathcal{M}_q$ with the (i,j)-th element:

$$\Big\langle \mathbb{E}\big[\nabla_{\theta_i} \boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) \exp(X\cdot)\big],\ \mathbb{E}\big[\nabla_{\theta_j} \boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) \exp(X\cdot)\big] \Big\rangle,$$

and

$$\sum_{k=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \mathrm{Re}\left(\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{\exp,k,T}\big(\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}})\big)' \sqrt{T}\, \overline{\boldsymbol{m}}_{\exp,k,T}(\boldsymbol{\theta}_o)\right) = \sum_{k=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{\exp,k}(\boldsymbol{\theta}_o)' \sqrt{T}\, \overline{\boldsymbol{m}}_{\exp,k,T}(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1). \tag{18}$$

In this case, (16) becomes

$$\sqrt{T}\big(\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}}) - \boldsymbol{\theta}_o\big) = -\mathcal{M}_q^{-1} \left[\sum_{k=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{\exp,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \sqrt{T}\, \overline{\boldsymbol{m}}_{\exp,\boldsymbol{k},T}(\boldsymbol{\theta}_o)\right] + o_{\mathbb{P}}(1), \tag{19}$$

which also has a limiting normal distribution. The result below is analogous to Theorem 3.7.

Theorem 3.8 *Given* [A1]–[A3] *and* [A5]–[A8], *if* $\mathcal{K}_T = o(T^{1/2})$, *then*

$$\sqrt{T}\big(\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_T) - \boldsymbol{\theta}_o\big) \xrightarrow{D} \mathcal{N}(0, \mathcal{V}),$$

where $\mathcal{V} = \mathcal{M}_q^{-1} \boldsymbol{\Omega}_q \mathcal{M}_q^{-1}$ *and* $\boldsymbol{\Omega}_q$ *is a* $q \times q$ *matrix with the* (i,j)*-th element:*

$$\Big\langle \mathbb{E}\left[\nabla_{\theta_i} \boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) \exp(X\cdot)\right],\ \mathbb{K}\,\mathbb{E}\left[\nabla_{\theta_j} \boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) \exp(X\cdot)\right] \Big\rangle.$$

For estimation of $\mathcal{V}$ in Theorem 3.8, note from (17) that $\mathcal{M}_q$ can be consistently estimated by

$$\sum_{k=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{\exp,k,T}\big(\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}})\big)' \nabla_{\boldsymbol{\theta}} \overline{\boldsymbol{m}}_{exp,k,T}\big(\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}})\big).$$

From [A8] and (18), $\mathbf{\Omega}_q$ can be consistently estimated by

$$\sum_{k=-\mathcal{K}_{\rm T}}^{\mathcal{K}_{\rm T}} \Big[\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{\exp,k,T}\big(\widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_{\rm T})\big)'\Big] \times \left[\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{h}\big(\boldsymbol{y}_t, \widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_{\rm T})\big)\varphi^2_{\exp,k}(x_t)\boldsymbol{h}\big(\boldsymbol{y}_t, \widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_{\rm T})\big)'\right] \times \Big[\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{\exp,k,T}\big(\widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_{\rm T})\big)\Big].$$

A consistent estimator of $\mathcal{V}$ is then readily constructed from these two estimators.

3.3 The Results for Multivariate $\boldsymbol{X}$

We now extend the asymptotic properties above to the case with multivariate $\boldsymbol{X}$. Recall that $\boldsymbol{X}$ is an $m \times 1$ vector of conditioning variables, and set $\mathcal{T} = [-\pi,\pi]^{m+1}$. Then the proposed instruments resulted from G now are

$$\varphi_{G,\boldsymbol{k}}(X) = \int_{[-\pi,\pi]^{m+1}} G\big(A(\boldsymbol{X},\boldsymbol{\tau})\big) \exp\left(-i\boldsymbol{k}'\boldsymbol{\tau}\right) \mathrm{d}\boldsymbol{\tau},$$

where $\boldsymbol{k} = (k_0, k_1, \ldots, k_m)'$. The required conditions of establishing asymptotics are unchanged, except for condition [A4]. We need [A4'] below in establishing asymptotics instead.

[A4'] G is real analytic but not a polynomial such that, w.p.1,

$$\sup_{\boldsymbol{\tau}\in\mathcal{T}} \left| \frac{\partial^j G(A(X,\boldsymbol{\tau}))}{\prod_{i=0}^m (\partial \tau_i)^{l_i}} \right| < \infty,$$

where $i = 0, 1, \ldots, m$, $j = 1, \ldots, m$, and $l_i = 0, 1, \ldots, j$ such that $\sum_{i=1}^m l_i = j$.

Besides, we now define $c(k_i) = |k_i|$ for $k_i \neq 0$ and $c(k_i) = 1$ for $k_i = 0$, $i = 0, 1, \ldots, m$. Similar to what proceeded in the proof of Lemma 3.1, a new bound on $\varphi_{G,\boldsymbol{k}}(\boldsymbol{X})$ when $\boldsymbol{X}$ is multivariate can be easily provided as

Lemma 3.9 *Given* [A4'], $|\varphi_{G,\boldsymbol{k}}(X)| \leq \Delta/[\prod_{i=0}^m c(k_i)]$ *w.p.1, where* Δ *is a real number.*

Given this result and the procedures analogous to the proofs of Theorem 3.3 and Theorem 3.7, we will have[5]

Theorem 3.10 *Given* [A1]–[A3], *and* [A4'], *if* $\mathcal{K}_T = o(T^{1/(m+1)})$, *then* $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_o$ *as* $T \to \infty$.

[5] Note that the dimension m affects the order of $\mathcal{K}_{\rm T}$ only through the implication rule and the generalized Chebyshev inequality in the proofs.

Theorem 3.11 *Given* [A1]–[A3], [A4'], *and* [A5]–[A8], *if* $\mathcal{K}_T = o(T^{1/(2m+2)})$, *then*

$$\sqrt{T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T) - \boldsymbol{\theta}_o\big) \xrightarrow{D} \mathcal{N}(0,\mathcal{V}),$$

where $\mathcal{V} = \mathcal{M}_q^{-1}\boldsymbol{\Omega}_q\mathcal{M}_q^{-1}$ *and* $\boldsymbol{\Omega}_q$ *is a* $q \times q$ *matrix with the* (i,j)*-th element:*

$$\left\langle \mathbb{E}\left[\nabla_{\theta_i}\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o)G(A(X,\cdot))\right],\ \mathbb{K}\,\mathbb{E}\left[\nabla_{\theta_j}\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o)G(A(X,\cdot))\right]\right\rangle.$$

4 Simulations

We now evaluate the finite-sample performance of the proposed estimator $\widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_{\mathrm{T}})$ and compare its performance with the nonlinear least squares (NLS) estimator:

$$\widehat{\boldsymbol{\theta}}_{\mathrm{NLS}} = \operatorname*{argmin}_{\boldsymbol{\theta}\in\Theta} \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta})^2,$$

and the estimator of Domínguez and Lobato (2004), $\widehat{\boldsymbol{\theta}}_{\mathrm{DL}}$ in (5). When a random variable is unbounded, its data x_t are transformed using a logistic mapping: $\exp(x_t)/[1+\exp(x_t)]$, which yields values between 0 and 1. Our comparison is based on the bias, standard error (SE), and mean squared error (MSE) of these estimators. The parameter estimates are computed using the GAUSS optimization procedure, OPTMUM, with the BFGS algorithm; for each optimization, three initial values are randomly drawn from the standard normal distribution. For the proposed estimator, we set $\mathcal{K}_{\mathrm{T}} = 5$; the effect of different $\mathcal{K}_{\mathrm{T}}$ on the proposed estimator will be examined in Section 4.4. In all experiments, the samples are $T = 50, 100, 200$; the number of replications is 5000.

4.1 The Experiments in Domínguez and Lobato (2004)

Following Domínguez and Lobato (2004), we postulate a simple nonlinear model:

$$Y = \theta_o^2 X + \theta_o X^2 + \epsilon, \quad \epsilon \sim \mathcal{N}(0,1),$$

where $\theta_o = 5/4$ is the unique solution to the conditional moment restriction: $\mathbb{E}(\epsilon|X) = 0$. We consider two cases: $X \sim \mathcal{N}(0,1)$ and $X \sim \mathcal{N}(1,1)$. In the former case, $\theta_o = 5/4$ is the only real solution to the unconditional moment restriction resulted from the "feasible" optimal instrument $(2\theta X + X^2)$; the other two solutions are complex: $-0.625 \pm 1.0533i$. When $X \sim \mathcal{N}(1,1)$, $\theta = -5/4$ and $\theta = -3$ also satisfy the unconditional moment restriction with the feasible optimal instrument. Yet, it can be shown that 5/4 is the global minimum of the MSE objective function, whereas the other two solutions are only local minima.[6] For comparison, our simulations here also includes the optimal instrument variable (OPIV) estimator:

$$\widehat{\theta}_{\text{OPIV}} = \operatorname*{argmin}_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^{T} (y_t - \theta^2 x_t - \theta x_t^2)(2\theta x_t + x_t^2) \right)^2,$$

which is different from the NLS estimator.[7]

The simulation results are summarized in Table 1. In both cases, the NLS estimator outperforms the other estimators in terms of bias, SE and MSE, while $\widehat{\theta}_{\text{OPIV}}$ has severe bias and large SE and is dominated by the other estimators. It can also be seen that the proposed estimator, $\widehat{\theta}(\exp, \mathcal{K}_{\text{T}})$, outperforms the Domínguez-Lobato estimator, $\widehat{\theta}_{\text{DL}}$, in terms of bias, SE and MSE for all samples when $X \sim \mathcal{N}(1,1)$. For the case $X \sim \mathcal{N}(0,1)$, the proposed estimator performs better than $\widehat{\theta}_{\text{DL}}$ for smaller samples ($T = 50$ and 100). Thus, the proposed estimator compares favorably with the Domínguez-Lobato estimator when there are multiple local minima. Note, however, that the NLS estimator need not always be the best estimator, as shown in Section 4.3.

4.2 Model with an Endogenous Regressor

We extend the previous experiment to the case that there is an endogenous regressor. The model specification is:

$$Y = \theta_o^2 Z + \theta_o Z^2 + \epsilon,$$

and $Z = X + \nu$, with

$$\begin{bmatrix} \epsilon \\ \nu \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

where $\theta_o = 5/4$, $\rho = 0.01, 0.1, 0.3, 0.5, 0.7, 0.9$, and $X \sim \mathcal{N}(0,1)$ is independent of ϵ and ν. Given this specification, $\mathbb{E}(\epsilon|X) = 0$. The simulation results are collected in Table 2.

[6] Domínguez and Lobato (2004, p. 1602) claimed that θ_o can not be globally identified by $\mathbb{E}[(Y - \theta^2 X - \theta X^2)(2\theta X + X^2)] = 0$, which is the first order condition of MSE minimization. This is not true because $\theta_o = 5/4$ is the global minimum, whereas the other solutions only lead to local minima.

[7] Domínguez and Lobato (2004, p. 1608) mentioned that the GMM estimator with the optimal instrument is also the NLS estimator. Yet, it can be seen that these two estimators would be different if the GMM estimator is defined as the OPIV estimator given above.

As expected, the biases of all three (NLS, Domínguez-Lobato and proposed) estimators become larger when ρ increases. In particular, the large biases of NLS estimator are induced by the inconsistency (due to the endogenous regressor), and such biases do not die out even when the sample size is large. Nonetheless, the proposed estimator is consistent and performs remarkably well. It performs significantly better than $\widehat{\theta}_{\mathrm{DL}}$ in terms of bias, SE and MSE for any ρ and any sample size. It also has much smaller bias than the NLS estimator, except only when $\rho = 0.01$ and $T = 100$. Although the NLS estimator typically has a smaller SE, the proposed estimator yields smaller MSE as long as the correlation between ϵ and ν is not too small (e.g., $\rho \geq 0.3$).

4.3 Noisy Disturbances

We now examine the effect of the disturbance variance on the performance of various estimators. The model is again

$$Y = \theta_o^2 X + \theta_o X^2 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2),$$

where $\theta_o = 5/4$, X is the uniform random variable on $(-1, 1)$ and independent of ϵ, and σ^2 =0.1, 1, 4, 9, and 16. It can be verified that there are 3 solutions to the unconditional moment restriction resulted from the "feasible" optimal instrument $(2\theta X + X^2)$: $\theta = 5/4$ and $(-25 \pm \sqrt{145})/40$, where 5/4 is the global minimum.

The results are summarized in Table 3. In contrast with the results in Table 1, the NLS estimator is no longer the best estimator even when there is a unique global minimum and the regressor is exogenous. In all cases, the proposed estimator is the one with the smallest bias, while the Domínguez-Lobato estimator has the largest bias. In terms of MSE, the proposed estimator, in general, dominates the Domínguez-Lobato estimator and may also outperform the NLS estimator when σ^2 is not too large. For example, for the sample $T = 100$, the proposed estimator has smaller MSE than the NLS estimator when $\sigma^2 \leq 1$, and it has smaller MSE than the Domínguez-Lobato estimator when $\sigma^2 \leq 9$. For $T = 200$, the proposed estimator has smaller MSE than the NLS estimator when $\sigma^2 \leq 4$ but dominates the Domínguez-Lobato estimator in all cases.

4.4 The Proposed Estimator with Various $\mathcal{K}_{\mathrm{T}}$

We now examine the effect of $\mathcal{K}_{\mathrm{T}}$ on the performance of the proposed estimator. The model specification is the same as that in Section 4.2, where the regressor is endogenous.

We consider the cases that ρ equals $0.1, 0.5$ and 0.9, and the sample $T = 50$, 100 and 200. We simulate the Domínguez-Lobato estimator and $\hat{\theta}(\mathcal{K}_{\mathrm{T}})$ with $\mathcal{K}_{\mathrm{T}} = 1, 2, \ldots, 10, 15, 20$. We do not consider the NLS estimator because it performs poorly when regressor is endogenous. To save space, we report only the results for $\rho = 0.5$ and $\rho = 0.9$, each with $T = 100, 200$ in Tables 4 and 5. In addition to the bias, SE and MSE, we also report their percentage changes when $\mathcal{K}_{\mathrm{T}}$ increases. For instance, for $\rho = 0.9$ and $T = 100$, the bias decreases 0.96%, SE decreases 1.78%, and MSE decreases 3.5% when $\mathcal{K}_{\mathrm{T}}$ increases from 1 to 2.

These tables show that, when $\mathcal{K}_{\mathrm{T}}$ increases, the proposed estimator becomes more efficient (with a smaller SE), while its bias typically decreases.[8] The percentage changes of bias and SE are typically small. In most cases, such changes are less than 0.1% when $\mathcal{K}_{\mathrm{T}}$ is greater than 5 or 6. These results suggest that the first few Fourier coefficients indeed contain the most information for identifying θ_o. Further increase of $\mathcal{K}_{\mathrm{T}}$ can only result in marginal improvements on the bias and SE. Furthermore, the proposed estimator again dominates the Domínguez-Lobato estimator in terms of bias, SE and MSE in all cases.

5 Concluding Remarks

This paper is concerned with consistent estimation of conditional moment restrictions. We consider a continuum of unconditional moments that can identify the parameters of interest and construct an objective function from which a consistent GMM estimator can be obtained. Although the proposed estimator performs quite well in finite samples, we do not have to confine ourselves with GMM estimation. Our approach is readily applied to derive other consistent estimators, such as the empirical likelihood estimator. Our preliminary simulations show that the empirical likelihood estimator also outperforms the Domínguez-Lobato estimator in terms of bias, yet it does not perform as well as the estimator proposed in the paper.

This paper does not deal with estimator efficiency (directly) in the proposed procedure. We note that an efficient estimator may be obtained from the proposed estimator via an additional Newton-Raphson step, as in Newey (1990, 1993) and Domínguez and Lobato (2004). In particular, an efficient estimator can be computed as:

$$\widehat{\boldsymbol{\theta}}_T^e = \widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}}) - \left[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} Q_T\big(\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}}), \mathcal{K}_{\mathrm{T}}\big)\right]^{-1} \nabla_{\boldsymbol{\theta}} Q_T\big(\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}}), \mathcal{K}_{\mathrm{T}}\big),$$

[8] In the case that $\rho = 0.5$ and $T = 100$, the bias of the proposed estimator increases but with a decreasing rate. This ill behavior may be due to the initial values generated in the simulations.

where $Q_T(\boldsymbol{\theta}, \mathcal{K}_{\mathrm{T}})$ is the objective function for the efficient estimator that can locally identify $\boldsymbol{\theta}_o$, and $\nabla_{\boldsymbol{\theta}} Q_T(\boldsymbol{\theta}, \mathcal{K}_{\mathrm{T}})$ and $\nabla_{\boldsymbol{\theta\theta'}} Q_T(\boldsymbol{\theta}, \mathcal{K}_{\mathrm{T}})$ are its gradient vector and Hessian matrix, respectively. Yet, it is practically difficult to estimate the gradient and Hessian matrix (e.g., Newey, 1990, 1993). Carrasco and Florens (2000) discussed an efficient estimation method based on projection along the estimated eigenfunctions of the covariance operator $\mathbb{K}$ in Theorem 3.7, which is quite cumbersome to implement. Finding an efficient and convenient estimator remains an important research direction.

Appendix

Proof of Lemma 3.1: Let Δ be a generic constant whose value varies in different cases. Recall that $\boldsymbol{X}$ is univariate, the affine transformation $A(X, \tau) = \tau_0 + \tau_1 X$ and

$$\begin{aligned}\varphi_{G,\boldsymbol{k}}(X) &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} G(\tau_0 + \tau_1 X)\exp(-ik_0\tau_0)\exp(-ik_1\tau_1)d\tau_0 \ \mathrm{d}\tau_1 \\ &= \int_{-\pi}^{\pi}\left[\int_{-\pi}^{\pi} G(\tau_0 + \tau_1 X)\exp(-ik_0\tau_0)d\tau_0\right]\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1.\end{aligned}$$

Applying integration by parts, for $k_0, k_1 \neq 0$, the term in the square bracket above can be expressed as

$$\begin{aligned}&\int_{-\pi}^{\pi} G(\tau_0 + \tau_1 X)\exp(-ik_0\tau_0)\,\mathrm{d}\tau_0 \\ &= \frac{i}{k_0}\left\{(-1)^{k_0}\left[G(\pi + \tau_1 X) - G(-\pi + \tau_1 X)\right] - \int_{-\pi}^{\pi} G_0(\tau_0 + \tau_1 X)\exp(-ik_0\tau_0)\,\mathrm{d}\tau_0\right\} \\ &:= \frac{i}{k_0}\left[Q_1(\boldsymbol{\tau}) - Q_2(\boldsymbol{\tau})\right],\end{aligned}$$

where $Q_1(\boldsymbol{\tau})$ denotes the first term and $Q_2(\boldsymbol{\tau})$ denotes the second one in above brace. It immediately follows that

$$\varphi_{G,\boldsymbol{k}}(X) = \frac{i}{k_0}\int_{-\pi}^{\pi}\left[Q_1(\boldsymbol{\tau}) - Q_2(\boldsymbol{\tau})\right]\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1,$$

and

$$\left|\varphi_{G,\boldsymbol{k}}(X)\right| \leq \frac{1}{|k_0|}\left\{\left|\int_{-\pi}^{\pi} Q_1(\boldsymbol{\tau})\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1\right| + \left|\int_{-\pi}^{\pi} Q_2(\boldsymbol{\tau})\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1\right|\right\}.$$

Again by integration by parts, we have

$$\begin{aligned}&\int_{-\pi}^{\pi} Q_1(\boldsymbol{\tau})\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1 \\ &= \frac{(-1)^{k_0} i}{k_1}\Big\{(-1)^{k_1}\left[G(\pi + \pi X) - G(-\pi + \pi X) - G(\pi - \pi X) + G(-\pi - \pi X)\right] \\ &\qquad - \int_{-\pi}^{\pi}\left[G_1(\pi + \tau_1 X) - G_1(-\pi + \tau_1 X)\right]\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1\Big\},\end{aligned}$$

and

$$\begin{aligned}&\int_{-\pi}^{\pi} Q_2(\boldsymbol{\tau})\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1 \\ &= \frac{i}{k_1}\Big\{(-1)^{k_1}\int_{-\pi}^{\pi}\left[G_0(\tau_0 + \pi X) - G_0(\tau_0 - \pi X)\right]\exp(-ik_0\tau_0)\,\mathrm{d}\tau_0 \\ &\qquad - \int_{-\pi}^{\pi}\left(\int_{\pi}^{\pi} G_{01}(\tau_0 + \tau_1 X)\exp(-ik_0\tau_0)\,\mathrm{d}\tau_0\right)\exp(-ik_1\tau_1)\,\mathrm{d}\tau_1\Big\}.\end{aligned}$$

Given [A4], it is easy to derive below inequalities :

$$
\begin{aligned}
&\left|\int_{-\pi}^{\pi} Q_1(\boldsymbol{\tau}) \exp(-ik_1\tau_1)\, \mathrm{d}\tau_1\right| \\
&\leq \frac{1}{|k_1|}\left[4 \sup_{\boldsymbol{\tau}\in\mathcal{T}} \left|G(\tau_0+\tau_1 X)\right| + 2\int_{-\pi}^{\pi} \sup_{\boldsymbol{\tau}\in\mathcal{T}} \left|G_1(\tau_0+\tau_1 X)\right| \left|\exp(-ik_1\tau_1)\right| \mathrm{d}\tau_1\right] \\
&\leq \frac{\Delta}{|k_1|},
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\int_{-\pi}^{\pi} Q_2(\boldsymbol{\tau}) \exp(-ik_1\tau_1)\, \mathrm{d}\tau_1\right| \\
&\leq \frac{1}{|k_1|}\left[2\int_{-\pi}^{\pi} \sup_{\boldsymbol{\tau}\in\mathcal{T}} \left|G_0(\tau_0+\tau_1 X)\right| \left|\exp(-ik_0\tau_0)\right| \mathrm{d}\tau_0 \right. \\
&\qquad \left. + \int_{-\pi}^{\pi} \left(\int_{\pi}^{\pi} \sup_{\boldsymbol{\tau}\in\mathcal{T}} \left|G_{01}(\tau_0+\tau_1 X)\right| \left|\exp(-ik_0\tau_0)\right| \mathrm{d}\tau_0\right) \left|\exp(-ik_1\tau_1)\right| \mathrm{d}\tau_1 \Big\}\right] \\
&\leq \frac{\Delta}{|k_1|}.
\end{aligned}
$$

Putting these results together, we have $|\varphi_{G,\boldsymbol{k}}(X)| \leq \Delta/(|k_0||k_1|)$ for $k_0, k_1 \neq 0$.

We can proceed as above and show that for $k_0 = 0$ and $k_1 \neq 0$, $|\varphi_{G,\boldsymbol{k}}(X)| \leq \Delta/|k_1|$ (for $k_0 \neq 0$ and $k_1 = 0$, $|\varphi_{G,\boldsymbol{k}}(X)| \leq \Delta/|k_0|$). Also, it is clear that $|\varphi_{G,0}(X)| \leq \Delta$. The proof is thus complete. □

Proof of Lemma 3.2: Let Δ again denote a generic constant whose value varies in different cases. Define

$$
\boldsymbol{\eta}_{G,\boldsymbol{k},t} = \boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(x_t) - \mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\varphi_{G,\boldsymbol{k}}(X)\big],
$$

for $t = 1,\ldots,T$ and $\boldsymbol{k} = (k_0,k_1)'$. By Lemma 3.1, $|\varphi_{G,\boldsymbol{k}}(X)| \leq \Delta/[c(k_0)c(k_1)]$. With [A3], we have

$$
\mathbb{E}\big[|\boldsymbol{\eta}_{G,\boldsymbol{k},t}|^2\big] \leq \mathbb{E}\big[|\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})|^2\, |\varphi_{G,\boldsymbol{k}}(X)|^2\big] \leq \frac{\Delta}{c(k_0)^2 c(k_1)^2}.
$$

Under [A1], these bounds lead to

$$\sum_{k_0,k_1=-\mathcal{K}_\mathrm{T}}^{\mathcal{K}_\mathrm{T}} \mathbb{E}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_{G,\boldsymbol{k},t}\right|^2\right]$$
$$= \frac{1}{T^2}\sum_{k_0,k_1=-\mathcal{K}_\mathrm{T}}^{\mathcal{K}_\mathrm{T}}\sum_{t=1}^{T}\mathbb{E}\left[\left|\boldsymbol{\eta}_{G,\boldsymbol{k},t}\right|^2\right]$$
$$\leq \frac{4\Delta}{T}\sum_{k_0=1}^{\mathcal{K}_\mathrm{T}}\frac{1}{k_0^2}\sum_{k_1=1}^{\mathcal{K}_\mathrm{T}}\frac{1}{k_1^2} + \frac{2\Delta}{T}\sum_{k_0=1}^{\mathcal{K}_\mathrm{T}}\frac{1}{k_0^2} + \frac{2\Delta}{T}\sum_{k_1=1}^{\mathcal{K}_\mathrm{T}}\frac{1}{k_1^2} + \frac{\Delta}{T}$$
$$\leq \frac{\Delta}{T},$$

by the fact that $\sum_{k=1}^{n} k^{-2} \leq 2 - 1/n \leq 2$. It follows from the implication rule and the generalized Chebyshev inequality that

$$\mathbb{P}\left[\sum_{k_0,k_1=-\mathcal{K}_\mathrm{T}}^{\mathcal{K}_\mathrm{T}}\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_{G,\boldsymbol{k},t}\right|^2 \geq \varepsilon\right]$$
$$\leq \sum_{k_0,k_1=-\mathcal{K}_\mathrm{T}}^{\mathcal{K}_\mathrm{T}}\mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_{G,\boldsymbol{k},t}\right|^2 \geq \frac{\varepsilon}{(2\mathcal{K}_\mathrm{T}+1)^2}\right]$$
$$\leq \frac{(2\mathcal{K}_\mathrm{T}+1)^2}{\varepsilon}\sum_{k_0,k_1=-\mathcal{K}_\mathrm{T}}^{\mathcal{K}_\mathrm{T}}\mathbb{E}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_{G,\boldsymbol{k},t}\right|^2\right]$$
$$\leq \frac{(2\mathcal{K}_\mathrm{T}+1)^2}{\varepsilon}\frac{\Delta}{T},$$

which holds uniformly in $\boldsymbol{\theta}$ because Δ does not depend on $\boldsymbol{\theta}$. It is then clear that this this bound can be made arbitrarily small when $\mathcal{K}_\mathrm{T} = o(T^{1/2})$. □

Proof of Theorem 3.3: The proposed estimator, $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_\mathrm{T})$, is the solution to the left-hand side of (13). Hence, it must converge to the unique minimizer, $\boldsymbol{\theta}_o$, of the right-hand side of (13) by Theorem 2.1 of Newey and McFadden (1994). □

Proof of Thorem 3.4: Given $G(A(X,\boldsymbol{\tau})) = \exp(X\tau)$, we have from the text that (14) holds when $\mathcal{K}_\mathrm{T} = o(T)$. Analogous to (13), we obtain

$$\sum_{k=-\mathcal{K}_T}^{\mathcal{K}_T}\left|\boldsymbol{m}_{\exp,k,T}(\boldsymbol{\theta})\right|^2 \xrightarrow{\mathbb{P}} \sum_{k=-\infty}^{\infty}\left|C_{\exp,k}(\boldsymbol{\theta})\right|^2,$$

uniformly in $\boldsymbol{\theta}$. The assertion again follows from Theorem 2.1 of Newey and McFadden (1994). □

Proof of Lemma 3.5: Given [A1]–[A4] and $\mathcal{K}_{\mathrm{T}} = o(T^{1/4})$, $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}}) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_o$. Hence, $\boldsymbol{\theta}_T^{\dagger} \to \boldsymbol{\theta}_o$. With [A6], we can apply a standard argument to get

$$\begin{aligned}\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{G,\boldsymbol{k},T}\big(\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})\big) - \nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) &\xrightarrow{\mathbb{P}} \mathbf{0},\\ \nabla_{\boldsymbol{\theta}} \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_T^{\dagger}) - \nabla_{\boldsymbol{\theta}} \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) &\xrightarrow{\mathbb{P}} \mathbf{0}.\end{aligned}$$

Also note that $\nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \nabla_{\boldsymbol{\theta}} \overline{C}_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)$ is real and

$$\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \nabla_{\boldsymbol{\theta}} \overline{C}_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o) \to \sum_{k_0,k_1=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \nabla_{\boldsymbol{\theta}} \overline{C}_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o).$$

Therefore, it suffices to show

$$\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \big(\nabla_{\boldsymbol{\theta}} \boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o)' \nabla_{\boldsymbol{\theta}} \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) - \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)' \nabla_{\boldsymbol{\theta}} \overline{C}_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)\big) \xrightarrow{\mathbb{P}} \mathbf{0}.$$

We shall show this convergence holds elementwise. For notation simplicity, we drop the subscript G and the argument $\boldsymbol{\theta}_o$ and write $\eta_{i,\boldsymbol{k}} = \nabla_{\theta_i} \boldsymbol{m}_{\boldsymbol{k},T} - \mathbb{E}[\nabla_{\theta_i} \boldsymbol{m}_{\boldsymbol{k},T}]$. The (i,j)-th element of the matrix above can be expressed as $\eta_{i,\boldsymbol{k}}' \nabla_{\theta_j} \overline{\boldsymbol{m}}_{\boldsymbol{k},T} + \nabla_{\theta_i} C_{\boldsymbol{k}}' \bar{\eta}_{j,\boldsymbol{k}}$. We need to show

$$\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \Big(\eta_{i,\boldsymbol{k}}' \nabla_{\theta_j} \overline{\boldsymbol{m}}_{\boldsymbol{k},T} + \nabla_{\theta_i} C_{\boldsymbol{k}}' \bar{\eta}_{j,\boldsymbol{k}}\Big) \xrightarrow{\mathbb{P}} 0.$$

Again by the implication rule and the generalized Chebyshev inequality, we have

$$\begin{aligned}&\mathbb{P}\left\{\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \left|\eta_{i,\boldsymbol{k}}' \nabla_{\theta_j} \overline{\boldsymbol{m}}_{\boldsymbol{k},T} + \nabla_{\theta_i} C_{\boldsymbol{k}}' \bar{\eta}_{j,\boldsymbol{k}}\right| \geq \epsilon\right\}\\ &\leq \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \mathbb{P}\left\{\left|\eta_{i,\boldsymbol{k}}' \nabla_{\theta_j} \overline{\boldsymbol{m}}_{\boldsymbol{k},T} + \nabla_{\theta_i} C_{\boldsymbol{k}}' \bar{\eta}_{j,\boldsymbol{k}}\right| \geq \frac{\epsilon}{(2\mathcal{K}_{\mathrm{T}}+1)^2}\right\}\\ &\leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon} \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \mathbb{E}\left[\left|\eta_{i,\boldsymbol{k}}' \nabla_{\theta_j} \overline{\boldsymbol{m}}_{\boldsymbol{k},T} + \nabla_{\theta_i} C_{\boldsymbol{k}}' \bar{\eta}_{j,\boldsymbol{k}}\right|\right]\\ &\leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon} \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} [\mathbb{E}\,|\eta_{i,\boldsymbol{k}}|^2]^{1/2} [\mathbb{E}\,|\nabla_{\theta_j} \overline{\boldsymbol{m}}_{\boldsymbol{k},T}|^2]^{1/2} + [\mathbb{E}\,|\nabla_{\theta_i} C_{\boldsymbol{k}}|^2]^{1/2} [\mathbb{E}\,|\bar{\eta}_{j,\boldsymbol{k}}|^2]^{1/2}.\end{aligned}$$

By [A1], [A6] and Lemma 3.1,

$$\mathbb{E}\,|\nabla_{\theta_j}\boldsymbol{m}_{\boldsymbol{k},T}|^2 = \frac{1}{T}\,\mathbb{E}\,|\nabla_{\theta_j}\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta})\varphi_{\boldsymbol{k}}(X)|^2 \leq \frac{\Delta}{Tc(k_0)^2c(k_1)^2}.$$

Similarly, $|\nabla_{\theta_i}C_{\boldsymbol{k}}|^2 \leq \Delta/[c(k_0)^2c(k_1)^2]$, and

$$\mathbb{E}\,|\eta_{i,\boldsymbol{k}}|^2 = \mathbb{E}\,|\nabla_{\theta_i}\boldsymbol{m}_{\boldsymbol{k},T}|^2 - \mathbb{E}\,|\nabla_{\theta_i}C_{\boldsymbol{k}}|^2 \leq \mathbb{E}\,|\nabla_{\theta_i}\boldsymbol{m}_{\boldsymbol{k},T}|^2 \leq \frac{\Delta}{Tc(k_0)^2c(k_1)^2}.$$

Putting these results together we have, similar to the proof of Lemma 3.2,

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}}\left|\eta_{i,\boldsymbol{k}}'\nabla_{\theta_j}\overline{\boldsymbol{m}}_{\boldsymbol{k},T} + \nabla_{\theta_i}C_{\boldsymbol{k}}'\,\bar{\eta}_{j,\boldsymbol{k}}\right| \geq \epsilon\right\}\\
&\quad\leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon}\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}}\left(\frac{\Delta}{Tc(k_0)^2c(k_1)^2} + \frac{\Delta}{\sqrt{T}c(k_0)^2c(k_1)^2}\right)\\
&\quad\leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon}\frac{\Delta}{\sqrt{T}},
\end{aligned}$$

which can be made arbitrarily small when $\mathcal{K}_{\mathrm{T}} = o(T^{1/4})$. □

Proof of Lemma 3.6: Similar to the proof of Lemma 3.5, given [A1]–[A6] and $\mathcal{K}_{\mathrm{T}} = o(T^{1/4})$, $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}}) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_o$, it is thus sufficient to show

$$\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}}\left[\nabla_{\boldsymbol{\theta}}\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) - \nabla_{\boldsymbol{\theta}}C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)\right]'\sqrt{T}\,\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \xrightarrow{\mathbb{P}} 0,$$

since

$$\nabla_{\boldsymbol{\theta}}\boldsymbol{m}_{G,\boldsymbol{k},T}\big(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_{\mathrm{T}})\big) - \nabla_{\boldsymbol{\theta}}\boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \xrightarrow{\mathbb{P}} \mathbf{0}$$

and

$$\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}}\nabla_{\boldsymbol{\theta}}C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)'\sqrt{T}\,\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \to \sum_{k_0,k_1=-\infty}^{\infty}\nabla_{\boldsymbol{\theta}}C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)'\sqrt{T}\,\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o),$$

where, by invoking the multiplication theorem,

$$\begin{aligned}
&\sum_{k_0,k_1=-\infty}^{\infty}\nabla_{\boldsymbol{\theta}}C_{G,\boldsymbol{k}}(\boldsymbol{\theta}_o)'\sqrt{T}\,\overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o)\\
&\qquad= \left\langle \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o)G\big(A(\boldsymbol{X},\cdot)\big)\right], \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta}_o)G\big(A(\boldsymbol{x}_t,\cdot)\big)\right\rangle
\end{aligned}$$

is real. Again let $\eta_{i,\boldsymbol{k}} = \nabla_{\theta_i} \boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) - \mathbb{E}[\nabla_{\theta_i} \boldsymbol{m}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o)]$ and by the implication rule and the generalized Chebyshev inequality, we have

$$\begin{aligned}
&\mathbb{P}\left\{ \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \left| \eta'_{i,\boldsymbol{k}} \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \right| \geq \epsilon \right\} \\
&\leq \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \mathbb{P}\left\{ \left| \eta'_{i,\boldsymbol{k}} \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \right| \geq \frac{\epsilon}{(2\mathcal{K}_{\mathrm{T}}+1)^2} \right\} \\
&\leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon} \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \mathbb{E}\left[\left| \eta'_{i,\boldsymbol{k}} \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \right| \right] \\
&\leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon} \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} [\mathbb{E}\, |\eta_{i,\boldsymbol{k}}|^2]^{1/2} \left[\mathbb{E} \left| \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \right|^2 \right]^{1/2} \\
&\leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon} \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} [\mathbb{E}\, |\eta_{i,\boldsymbol{k}}|^2]^{1/2} \left[\mathbb{E}\, |\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) \varphi_{G,\boldsymbol{k}}(X)|^2 \right]^{1/2},
\end{aligned}$$

where the last inequality, given [A1], comes form the fact that

$$\mathbb{E} \left| \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \right|^2 = \mathbb{E} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t, \boldsymbol{\theta}_o) \varphi_{G,\boldsymbol{k}}(x_t) \right|^2 = \mathbb{E} \left| \boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) \varphi_{G,\boldsymbol{k}}(X) \right|^2.$$

Since we already have, form the proof of Lemma 3.5, that

$$\mathbb{E}\, |\eta_{i,\boldsymbol{k}}|^2 \leq \frac{\Delta}{T c(k_0)^2 c(k_1)^2},$$

and

$$\mathbb{E}\, |\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{\theta}_o) \varphi_{G,\boldsymbol{k}}(X)|^2 \leq \frac{\Delta}{T c(k_0)^2 c(k_1)^2},$$

it follows that

$$\mathbb{P}\left\{ \sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \left| \eta'_{i,\boldsymbol{k}} \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \right| \geq \epsilon \right\} \leq \frac{(2\mathcal{K}_{\mathrm{T}}+1)^2}{\epsilon} \frac{\Delta}{\sqrt{T}},$$

which completes the proof when this bound can be arbitrarily small given $\mathcal{K}_{\mathrm{T}} = o(T^{1/4})$ and $T \to \infty$. □

Proof of Theorem 3.7: From [A8], we know $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t, \boldsymbol{\theta}_o) G(A(x_t, \cdot)) \xrightarrow{D} \mathbb{Z}$, where $\mathbb{Z}$ is a Gaussian random element in $L_2([-\pi, \pi]^2)$ with the covariance operator $\mathbb{K}$. By invoking the multiplication theorem, we have

$$
\begin{aligned}
&\sum_{k_0,k_1=-\mathcal{K}_{\mathrm{T}}}^{\mathcal{K}_{\mathrm{T}}} \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o)' \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) \\
&\quad = \sum_{k_0,k_1=-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} C_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o)' \sqrt{T}\, \overline{\boldsymbol{m}}_{G,\boldsymbol{k},T}(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1) \\
&\quad = \left(\left\langle \nabla_{\theta_i} \mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G\big(A(X,\cdot)\big)\big],\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta}_o) G(A(x_t,\cdot)) \right\rangle \right)_{i=1,\ldots,p} + o_{\mathbb{P}}(1) \\
&\quad = \left(\left\langle \nabla_{\theta_i} \mathbb{E}\big[\boldsymbol{h}(\boldsymbol{Y},\boldsymbol{\theta}_o) G\big(A(X,\cdot)\big)\big],\ \mathbb{Z} \right\rangle \right)_{j=1,\ldots,p} + o_{\mathbb{P}}(1) \\
&\quad \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_q).
\end{aligned}
$$

The conclusion now follows from (16). □

Proof of Theorem 3.8: In this case, [A8] ensures $T^{-1/2}\sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{y}_t,\boldsymbol{\theta}_o)\exp(x_t,\cdot) \xrightarrow{D} \mathbb{Z}$, where $\mathbb{Z}$ is a Gaussian random element in $L_2[-\pi,\pi]$ with the covariance operator $\mathbb{K}$. Analogous to the proof for Theorem 3.7, the conclusion follows from (19). □

References

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions, *Econometrica*, **71**, 1975–1843.

Bierens, H. J. (1982). Consistent model specification test *Journal of Econometrics*, **20**, 105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form, *Econometrica*, **58**, 1443–1458.

Bierens, H. J. (1994). *Topics in Advanced Econometrics*, New York: Cambridge University Press.

Chen, X. and H. White (1996). Laws of large numbers for Hilbert space-valued mixingales with applications, *Econometric Theory*, **12**, 284–304.

Chen, X. and H. White (1998). Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications, *Econometric Theory*, **14**, 260–284.

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

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics*, **34**, 305–334.

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

Donald, S. G., G. W. Imbens, and W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions, *Journal of Econometrics*, **117**, 55-93.

Gallant, A. R. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford, UK: Basil Blackwell.

Hansen, B. E. and K. D. West (2002), Generalized Method of Moments and Macroeconomics, *Journal of Business and Economic Statistics*, **20**, 460–469.

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

Hansen, L. P. and K. J. Singleton (1982). Generalized instrumental variable estimation of nonlinear rational expectations models, *Econometrica*, **50**, 1269–1286.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes, *Annals of Statistics*, **25**, 2084–2102.

Kitamura, Y., G. Tripathi, and H. Ahn (2004). Empirical likelihood-based inference in conditional moment restriction models, *Econometrica*, **72**, 1167–1714.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models, *Econometrica*, **58**, 809–837.

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions, in G. Maddala, C. Rao, and H. Vinod (eds.), *Handbook of Statistics*, **11**, pp. 2111–2245. Elsevier Science.

Qin, J. and J. Lawless (1994). Empirical likelihood and generalized estimating equations, *Annals of Statistics*, **22**, 300–325.

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

Stock, J. H. and J. H. Wright (2000). GMM with weak identification, *Econometrica*, **68**, 1055–1096.

Stuart, R. D. (1961). *An Introduction to Fourier Analysis*, New York: Halsted Press.

White, H. (1989). An additional hidden unit test for neglected nonlinearity, *Proceedings of the International Joint Conference on Neural Networks*, Vol. 2, pp. 451–455. New York: IEEE Press.

Table 1: Models in Domínguez and Lobato (2004) with exogenous regressors.

Sample		$X \sim \mathcal{N}(0,1)$			$X \sim \mathcal{N}(1,1)$		
T	Estimator	Bias	SE	MSE	Bias	SE	MSE
50	$\hat{\theta}_{\text{NLS}}$	-0.0006	0.0501	0.0025	-0.0083	0.1881	0.0354
	$\hat{\theta}_{\text{DL}}$	-0.0390	0.2282	0.0536	-0.0336	0.3667	0.1355
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	-0.0061	0.1600	0.0256	-0.0249	0.3308	0.1100
	$\hat{\theta}_{\text{OPIV}}$	-0.2222	0.6288	0.4447	-1.6922	1.2783	4.4972
100	$\hat{\theta}_{\text{NLS}}$	-0.0004	0.0342	0.0012	-0.0071	0.1713	0.0294
	$\hat{\theta}_{\text{DL}}$	-0.0152	0.1541	0.0240	-0.0316	0.3595	0.1302
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	-0.0059	0.1511	0.0228	-0.0217	0.3094	0.0962
	$\hat{\theta}_{\text{OPIV}}$	-0.1480	0.5096	0.2815	-1.7217	1.2619	4.5564
200	$\hat{\theta}_{\text{NLS}}$	-0.0004	0.0239	0.0006	-0.0025	0.1035	0.0107
	$\hat{\theta}_{\text{DL}}$	-0.0017	0.0864	0.0075	-0.0191	0.2796	0.0785
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	-0.0045	0.1390	0.0193	-0.0116	0.2278	0.0520
	$\hat{\theta}_{\text{OPIV}}$	-0.0931	0.3994	0.1681	-1.6649	1.2859	4.4250

Table 2: Models with an endogenous regressor.

		$T = 50$			$T = 100$			$T = 200$		
ρ	Est.	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.01	$\hat{\theta}_{\mathrm{NLS}}$	0.0009	0.0317	0.0010	0.0005	0.0212	0.0004	0.0011	0.0146	0.0002
	$\hat{\theta}_{\mathrm{DL}}$	−0.0103	0.1165	0.0137	−0.0062	0.0809	0.0066	−0.0027	0.0561	0.0032
	$\hat{\theta}(\mathcal{K}_{\mathrm{T}})$	−0.0003	0.0561	0.0031	−0.0009	0.0365	0.0013	0.0001	0.0245	0.0006
0.1	$\hat{\theta}_{\mathrm{NLS}}$	0.0097	0.0313	0.0011	0.0102	0.0210	0.0005	0.0103	0.0146	0.0003
	$\hat{\theta}_{\mathrm{DL}}$	−0.0116	0.1153	0.0134	−0.0069	0.0816	0.0067	−0.0036	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_{\mathrm{T}})$	−0.0021	0.0550	0.0030	−0.0010	0.0358	0.0013	−0.0006	0.0242	0.0006
0.3	$\hat{\theta}_{\mathrm{NLS}}$	0.0315	0.0310	0.0020	0.0311	0.0209	0.0014	0.0315	0.0144	0.0012
	$\hat{\theta}_{\mathrm{DL}}$	−0.0125	0.1214	0.0149	−0.0061	0.0819	0.0067	−0.0032	0.0585	0.0034
	$\hat{\theta}(\mathcal{K}_{\mathrm{T}})$	− 0.0039	0.0565	0.0032	−0.0016	0.0358	0.0013	−0.0002	0.0244	0.0006
0.5	$\hat{\theta}_{\mathrm{NLS}}$	0.0539	0.0311	0.0039	0.0527	0.0207	0.0032	0.0520	0.0143	0.0029
	$\hat{\theta}_{\mathrm{DL}}$	−0.0125	0.1231	0.0153	−0.0045	0.0817	0.0067	−0.0017	0.0570	0.0033
	$\hat{\theta}(\mathcal{K}_{\mathrm{T}})$	−0.0056	0.0596	0.0036	−0.0021	0.0366	0.0013	−0.0010	0.0247	0.0006
0.7	$\hat{\theta}_{\mathrm{NLS}}$	0.0746	0.0298	0.0064	0.0739	0.0196	0.0058	0.0731	0.0140	0.0055
	$\hat{\theta}_{\mathrm{DL}}$	−0.0153	0.1242	0.0156	−0.0083	0.0840	0.0071	−0.0053	0.0588	0.0035
	$\hat{\theta}(\mathcal{K}_{\mathrm{T}})$	−0.0097	0.0574	0.0034	−0.0038	0.0366	0.0014	−0.0020	0.0247	0.0006
0.9	$\hat{\theta}_{\mathrm{NLS}}$	0.0972	0.0285	0.0103	0.0953	0.0190	0.0094	0.0942	0.0134	0.0091
	$\hat{\theta}_{\mathrm{DL}}$	−0.0166	0.1288	0.0169	−0.0086	0.0845	0.0072	−0.0042	0.0598	0.0036
	$\hat{\theta}(\mathcal{K}_{\mathrm{T}})$	−0.0117	0.0947	0.0091	−0.0053	0.0370	0.0014	−0.0019	0.0250	0.0006

Table 3: Models with different disturbance variances.

σ^2	Est.	$T = 50$			$T = 100$			$T = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.1	$\hat{\theta}_{\text{NLS}}$	−0.2827	0.7836	0.6939	−0.2511	0.7392	0.6094	−0.2523	0.7433	0.6160
	$\hat{\theta}_{\text{DL}}$	−0.4645	0.8678	0.9687	−0.3938	0.8108	0.8124	−0.3701	0.7900	0.7610
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	−0.1129	0.5687	0.3361	−0.0491	0.4016	0.1637	−0.0317	0.3467	0.1212
1	$\hat{\theta}_{\text{NLS}}$	−0.5845	1.0572	1.4591	−0.4451	0.9481	1.0968	−0.3089	0.8158	0.7608
	$\hat{\theta}_{\text{DL}}$	−0.9491	1.0820	2.0711	−0.7899	1.0236	1.6715	−0.6692	0.9776	1.4033
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	−0.3478	1.1121	1.3574	−0.1724	0.8191	0.7005	−0.0880	0.5790	0.3429
4	$\hat{\theta}_{\text{NLS}}$	−0.8508	1.2109	2.1899	−0.7320	1.1470	1.8513	−0.5897	1.0582	1.4673
	$\hat{\theta}_{\text{DL}}$	−1.2441	1.1875	2.9576	−1.1126	1.1187	2.4891	−0.9874	1.0752	2.1307
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	−0.5081	1.5507	2.6623	−0.3767	1.3357	1.9257	−0.2599	1.0442	1.1577
9	$\hat{\theta}_{\text{NLS}}$	−0.9452	1.2880	2.5519	−0.8877	1.2281	2.2960	−0.7253	1.1491	1.8463
	$\hat{\theta}_{\text{DL}}$	−1.3698	1.2738	3.4985	−1.2821	1.1920	3.0644	−1.1359	1.1210	2.5467
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	−0.5013	1.8507	3.6759	−0.4600	1.6048	2.7864	−0.3355	1.3225	1.8612
16	$\hat{\theta}_{\text{NLS}}$	−1.0299	1.3672	2.9295	−0.9329	1.2814	2.5121	−0.8481	1.2134	2.1915
	$\hat{\theta}_{\text{DL}}$	−1.4794	1.3772	4.0848	−1.3361	1.2439	3.3321	−1.2542	1.1681	2.9371
	$\hat{\theta}(\mathcal{K}_{\text{T}})$	−0.5085	2.1182	4.7443	−0.3882	1.7852	3.3372	−0.3912	1.5377	2.5172

Table 4: The performance of $\hat{\theta}(\mathcal{K}_{\text{T}})$ with various $\mathcal{K}_{\text{T}}$: $\rho = 0.5$.

	$T = 100$					
$\mathcal{K}_{\text{T}}$	Bias	Bias(+%)	SE	SE(+%)	MSE	MSE(+%)
1	−0.00191	.	0.03708	.	0.00138	.
2	−0.00191	0.11425	0.03658	−1.34396	0.00134	−2.66222
3	−0.00191	0.05820	0.03638	−0.54672	0.00133	−1.08718
4	−0.00191	0.03550	0.03628	−0.29128	0.00132	−0.57993
5	−0.00191	0.02381	0.03621	−0.17973	0.00131	−0.35801
6	−0.00191	0.01703	0.03617	−0.12158	0.00131	−0.24224
7	−0.00191	0.01276	0.03614	−0.08758	0.00131	−0.17453
8	−0.00191	0.00991	0.03611	−0.06603	0.00131	−0.13160
9	−0.00191	0.00791	0.03609	−0.05154	0.00131	−0.10272
10	−0.00191	0.00646	0.03608	−0.04133	0.00131	−0.08238
15	−0.00191	0.02011	0.03603	−0.12436	0.00130	−0.24776
20	−0.00191	0.01047	0.03601	−0.06237	0.00130	−0.12428
$\hat{\theta}_{\text{DL}}$	−0.00552		0.08383		0.00706	
	$T = 200$					
$\mathcal{K}_{\text{T}}$	Bias	Bias(+%)	SE	SE(+%)	MSE	MSE(+%)
1	−0.00161	.	0.02545	.	0.00065	.
2	−0.00159	−1.35402	0.02509	−1.39866	0.00063	−2.77740
3	−0.00158	−0.57188	0.02495	−0.57062	0.00062	−1.13800
4	−0.00158	−0.30861	0.02487	−0.30440	0.00062	−0.60790
5	−0.00157	−0.19151	0.02483	−0.18794	0.00062	−0.37555
6	−0.00157	−0.12995	0.02480	−0.12719	0.00062	−0.25423
7	−0.00157	−0.09378	0.02477	−0.09164	0.00062	−0.18322
8	−0.00157	−0.07078	0.02476	−0.06911	0.00062	−0.13819
9	−0.00157	−0.05529	0.02474	−0.05395	0.00061	−0.10788
10	−0.00157	−0.04436	0.02473	−0.04327	0.00061	−0.08653
15	−0.00157	−0.13356	0.02470	−0.13021	0.00061	−0.26027
20	−0.00156	−0.06702	0.02468	−0.06531	0.00061	−0.13059
$\hat{\theta}_{\text{DL}}$	−0.00514		0.05945		0.00356	

Table 5: The performance of $\hat{\theta}(\mathcal{K}_{\mathrm{T}})$ with various $\mathcal{K}_{\mathrm{T}}$: $\rho = 0.9$.

	$T = 100$					
$\mathcal{K}_{\mathrm{T}}$	Bias	Bias(+%)	SE	SE(+%)	MSE	MSE(+%)
1	−0.00627	.	0.08508	.	0.00728	.
2	−0.00621	−0.95742	0.08356	−1.78127	0.00702	−3.52203
3	−0.00618	−0.37986	0.08299	−0.68917	0.00692	−1.37021
4	−0.00617	−0.19960	0.08269	−0.35995	0.00687	−0.71685
5	−0.00616	−0.12208	0.08251	−0.21985	0.00684	−0.43814
6	−0.00616	−0.08209	0.08238	−0.14783	0.00682	−0.29472
7	−0.00616	−0.05888	0.08230	−0.10608	0.00681	−0.21152
8	−0.00615	−0.04425	0.08223	−0.07977	0.00680	−0.15907
9	−0.00615	−0.03445	0.08218	−0.06213	0.00679	−0.12392
10	−0.00615	−0.02757	0.08214	−0.04975	0.00678	−0.09923
15	−0.00614	−0.08263	0.08202	−0.14930	0.00676	−0.29764
20	−0.00614	−0.04125	0.08196	−0.07468	0.00675	−0.14893
$\hat{\theta}_{\mathrm{DL}}$	−0.01039	.	0.08930	.	0.00808	.

	$T = 200$					
$\mathcal{K}_{\mathrm{T}}$	Bias	Bias(+%)	SE	SE(+%)	MSE	MSE(+%)
1	−0.00257	.	0.02498	.	0.00063	.
2	−0.00255	−0.48005	0.02464	−1.36135	0.00061	−2.68593
3	−0.00255	−0.18776	0.02451	−0.55611	0.00061	−1.10132
4	−0.00255	−0.09742	0.02443	−0.29676	0.00060	−0.58838
5	−0.00255	−0.05904	0.02439	−0.18324	0.00060	−0.36348
6	−0.00254	−0.03942	0.02436	−0.12401	0.00060	−0.24604
7	−0.00254	−0.02812	0.02434	−0.08935	0.00060	−0.17730
8	−0.00254	−0.02105	0.02432	−0.06738	0.00060	−0.13371
9	−0.00254	−0.01633	0.02431	−0.05259	0.00060	−0.10438
10	−0.00254	−0.01303	0.02430	−0.04218	0.00060	−0.08371
15	−0.00254	−0.03887	0.02427	−0.12692	0.00060	−0.25177
20	−0.00254	−0.01928	0.02425	−0.06365	0.00059	−0.12630
$\hat{\theta}_{\mathrm{DL}}$	−0.00480		0.05960		0.00357	