Journal of Econometrics 165 (2011) 87–99

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimation of conditional moment restrictions without assuming parameter identifiability in the implied unconditional moments☆

Shih-Hsun Hsu[a], Chung-Ming Kuan[b,*]

[a] Department of Economics, National Chengchi University, Taiwan

[b] Department of Finance, National Taiwan University, Taipei 106, Taiwan

ARTICLE INFO

Article history:
Available online 12 May 2011

JEL classification:
C12
C22

Keywords:
Conditional moment restrictions
Fourier coefficients
Generically comprehensive revealing function
Global identifiability
GMM

ABSTRACT

A well-known difficulty in estimating conditional moment restrictions is that the parameters of interest need not be globally identified by the implied unconditional moments. In this paper, we propose an approach to constructing a continuum of unconditional moments that can ensure parameter identifiability. These unconditional moments depend on the "instruments" generated from a "generically comprehensively revealing" function, and they are further projected along the exponential Fourier series. The objective function is based on the resulting Fourier coefficients, from which an estimator can be easily computed. 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. An efficient estimator is also readily obtained via the conventional two-step GMM method. 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.

1. Introduction

To estimate the parameters in conditional moment restrictions, it is typical to find a finite set of unconditional moment restrictions implied by the original 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). A leading example is the instrumental-variable estimation method for regression models. This approach hinges on the assumption that the parameters in the conditional restrictions can be globally identified by the implied, unconditional restrictions. With this assumption, estimator consistency is not really a problem and holds 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 et al. (2003).

Domínguez and Lobato (2004) challenge the assumption of global identifiability and show that the unconditional moments, when chosen arbitrarily, need not be equivalent to the original conditional restrictions. The identification problem may arise even when the unconditional moments are based on the so-called optimal instruments. Without assuming global identifiability, Domínguez and Lobato (2004) introduce the "instruments" generated from an indicator function and construct a continuum of unconditional moment restrictions that can identify the parameters of interest. However, there are some disadvantages of their method. 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 (Carrasco and Florens, 2000). Third, it is not easy to obtain an efficient estimate from their consistent estimate.

In this paper, we propose a different approach to constructing a continuum of unconditional moments that can ensure parameter identifiability. These unconditional moments depend on the "instruments" generated from the class of "generically comprehensively revealing" (GCR) functions (Stinchcombe and White, 1998), and these moments are further projected along the exponential Fourier series. The objective function is then based on the resulting Fourier coefficients, from which an estimator can be easily computed. A novel feature of our method is that it utilizes all possible information in the conditioning variables because *all* unconditional moments have been incorporated into each Fourier coefficient. Moreover, an efficient estimator can be obtained via the conventional two-step GMM method, which is computationally simpler than that of Carrasco and Florens (2000).

☆ This paper was originally entitled: "Consistent Parameter Estimation for Conditional Moment Restrictions."

* Corresponding author.
E-mail address: ckuan@ntu.edu.tw (C.-M. Kuan).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.05.008

We first show 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" generated from the exponential function, a special case in the class of GCR functions. For such instruments, the unconditional moments and Fourier coefficients have analytic forms, which greatly facilitate estimation in practice. Our simulations confirm that, under various settings, the proposed consistent and efficient estimators perform significantly better than that of Domínguez and Lobato (2004) in terms of bias, standard error, and mean squared error. The proposed estimators also outperform the estimator based on the optimal instruments.

This paper is organized as follows. We introduce the new class of consistent estimators in Section 2 and establish its consistency and asymptotic normality in Section 3. The Efficient estimator based on the proposed consistent estimator is discussed in Section 4. The simulation results are reported in Section 5. Section 6 concludes this paper. All proofs are deferred to Appendix.

2. Consistent estimation

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

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

where $\mathbf{h}$ is a $p \times 1$ vector of functions, $\mathbf{Y}$ is an $r \times 1$ vector of data variables, and $\mathbf{X}$ is an $m \times 1$ vector of conditioning variables. Without loss of generality, we shall work on the case that $\mathbf{X}$ 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:

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

for all measurable functions f, where each $f(\mathbf{X})$ may be interpreted as an "instrument" that helps us to identify $\boldsymbol{\theta}_o$. In practice, one typically forms an estimating function by subjectively choosing certain instruments, such as the square and cross products of the elements in $\mathbf{X}$. This would not be a problem in a linear model if the resulting unconditional moments can exactly identify θ_o. Yet, when $\mathbf{h}$ is nonlinear in $\boldsymbol{\theta}_o$, Domínguez and Lobato (2004) show 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 $\mathbf{X}$. 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 $\mathbf{X}$ (Bierens, 1982, 1990; Stinchcombe and White, 1998). Domínguez and Lobato (2004) set the instruments as $\mathbf{1}(\mathbf{X} \le \boldsymbol{\tau}) = \prod_{j=1}^{m} \mathbf{1}(X_j \le \tau_j)$, where $\mathbf{1}(A)$ is the indicator function of the event A. This leads to a continuum of unconditional moments indexed by $\boldsymbol{\tau}$ that are equivalent to (1):

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

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

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

with $P(\boldsymbol{\tau})$ a distribution function of $\boldsymbol{\tau}$ and $|\cdot|$ denotes the Euclidean norm.

A natural choice of $P(\boldsymbol{\tau})$ is $P_{\mathbf{X}}(\boldsymbol{\tau})$, the distribution function of $\mathbf{X}$. The L_2-norm in (4) can then be well approximated by the sample average. Domínguez and Lobato (2004) suggest the following estimator:

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

where $\mathbf{y}_t$ and $\mathbf{x}_t$ are the sample observations of $\mathbf{Y}$ and $\mathbf{X}$, respectively, and $\boldsymbol{\tau}_k = \mathbf{x}_k$, $k = 1, \ldots, T$. Clearly, this is a GMM estimator based on T unconditional moments induced by the indicator function.[2]

2.1. A class of consistent estimators

The indicator function is not the only choice for the desired instruments; Stinchcombe and White (1998) demonstrate that any GCR function will also do. Specifically, let $\Lambda_G(\mathcal{T})$ be the collection of $\lambda(\mathbf{X}) = G(A(\mathbf{X}, \boldsymbol{\tau}))$ with $\boldsymbol{\tau} \in \mathcal{T} \subset \mathbb{R}^{m+1}$ and $A(\mathbf{X}, \boldsymbol{\tau}) = \tau_0 + \sum_{j=1}^{m} X_j \tau_j$. Λ_G is said to be GCR if for all $\mathcal{T}$ with nonempty interior, the uniform closure of the span of $\Lambda_G(\mathcal{T})$ contains $C(\mathbf{B})$ for every compact $\mathbf{B}$, where $C(\mathbf{B})$ is the set of all bounded and continuous functions on $\mathbf{B}$. The function G is said to be GCR if Λ_G is GCR. Corollary 3.9 of Stinchcombe and White (1998) shows that a real analytic function is GCR if and only if it is not a polynomial (of a finite degree).[3] Note that polynomials of finite degree are not uniformly dense in the set of all continuous and bounded functions and hence cannot be a GCR function. The legitimate choices of G are, for example, the exponential function (Bierens, 1982, 1990) or the logistic function (White, 1989).

The discussion above suggests that, when G is GCR, (2) holds with $G(A(\mathbf{X}, \boldsymbol{\tau}))$ as legitimate instruments and $\boldsymbol{\tau}$ in an arbitrarily chosen index set $\mathcal{T}$ in $\mathbb{R}^{m+1}$. The unconditional moment restrictions induced by a GCR function are

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

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):

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

In contrast with Domínguez and Lobato (2004), there is no natural choice of $P(\boldsymbol{\tau})$, and it is 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 $\boldsymbol{\theta}_o$ without assuming parameter identifiability in the implied unconditional moments. This approach finds a condition equivalent to the L_2-norm in (7). To this end, we project the unconditional moments in (6) along the exponential Fourier series and obtain

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

[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.

[2] Alternatively, Ai and Chen (2003) and Kitamura et al. (2004) consider nonparametric estimation methods that deal with the conditional moments directly.

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

where $\mathcal{S} := \{\mathbf{k} = [k_0, k_1, \ldots, k_m]' \in \mathbb{Z}^{m+1}\}$ with $k_i = 0, \pm 1, \pm 2, \ldots, \pm\infty$, and $C_{G,\mathbf{k}}(\boldsymbol{\theta})$ is a Fourier coefficient:

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

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

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

in which the index parameter $\boldsymbol{\tau}$ has been integrated out.

We shall use the following notations. Given a complex number 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 numbers $\mathbf{f}$, its complex conjugate, real part and imaginary part are defined elementwise. Then, $|\mathbf{f}|^2 = \mathbf{f}'\bar{\mathbf{f}}$. Apart from a scaling factor, Parseval's Theorem implies that the L_2-norm in (7) is equivalent to

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

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

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

where the right-hand side no longer involves $\boldsymbol{\tau}$; cf. (7).

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

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

where $\mathcal{S}(\mathcal{K}_{\mathrm{T}})$ is a subset of $\mathcal{S}$ with $k_i = 0, \pm 1, \ldots, \pm\mathcal{K}_T$, such that $\mathcal{K}_{\mathrm{T}}$ grows with T but at a slower rate. The proposed estimator (10) depends on the function G, and it is also a GMM estimator based on $(2\mathcal{K}_{\mathrm{T}}+1)^{m+1}$ unconditional moments with the identity weighting matrix. Hence, $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})$ is not an efficient estimator in general.

Note that the Domínguez–Lobato estimator (5) relies only on a finite number of unconditional moments determined by the sample observations. By contrast, the proposed estimator (10) utilizes all possible information in estimation because each $\varphi_{G,\mathbf{k}}$ 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 eigenfunctions. Moreover, a regularization parameter must be determined in practice so as to ensure the invertibility of the estimated covariance operator.

2.2. A specific estimator

To compute the proposed estimator, we follow Bierens (1982, 1990) and 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 present the 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, $\exp(A(\mathbf{X}, \boldsymbol{\tau}))$ with $\boldsymbol{\tau} \in \mathbb{R}^{m+1}$ and $\exp(\mathbf{X}'\boldsymbol{\tau})$ 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 $\exp(\mathbf{X}'\boldsymbol{\tau})$ 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 $\mathcal{S} = \{\mathbf{k} = [k_1, \ldots, k_m]' \in \mathbb{Z}^m\}$.

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

$$\varphi_{\exp,\mathbf{k}}(\mathbf{X}) = \int_{\mathcal{T}} \exp(\mathbf{X}'\boldsymbol{\tau}) \exp(-\mathrm{i}\mathbf{k}'\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}$$
$$= \varphi_{\exp,k_1}(X_1)\cdots\varphi_{\exp,k_m}(X_m), \quad \mathbf{k} \in \mathcal{S}, \tag{11}$$

where

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

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

$$\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\mathrm{T}}) = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{\mathbf{k}\in\mathcal{S}(\mathcal{K}_{\mathrm{T}})} \left| \frac{1}{T}\sum_{t=1}^{T} \mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta})\varphi_{\exp,\mathbf{k}}(\mathbf{x}_t)\right|^2, \tag{12}$$

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

2.3. Implementing the proposed method

To summarize, the proposed consistent estimator for $\boldsymbol{\theta}_o$ in the conditional moment restriction (1) can be computed via the following steps.

1. Choose a GCR function G, a subset $\mathcal{T} \subset \mathbb{R}^{m+1}$ with a nonempty interior, and an integer $\mathcal{K}_{\mathrm{T}}$ smaller than T.
2. Denote $\mathcal{S}(\mathcal{K}_{\mathrm{T}}) := \{\mathbf{k} = [k_0, k_1, \ldots, k_m]', k_i = 0, \pm 1, \pm 2, \ldots, \pm\mathcal{K}_T\}$, and compute the instrument
$$\varphi_{G,\mathbf{k}}(\mathbf{X}) = \int_{\mathcal{T}} G(A(\mathbf{X}, \boldsymbol{\tau})) \exp(-\mathrm{i}\mathbf{k}'\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}$$
for $\mathbf{k} \in \mathcal{S}(\mathcal{K}_{\mathrm{T}})$. When G is the exponential function and $\mathcal{T} = [-\pi, \pi]^m$, $\varphi_{G,\mathbf{k}}(\mathbf{X})$ has the analytic form given in (11).
3. Using a GMM estimation program with the identity weighting matrix, the proposed estimator is computed as
$$\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}}) = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{\mathbf{k}\in\mathcal{S}(\mathcal{K}_{\mathrm{T}})} \left| \frac{1}{T}\sum_{t=1}^{T} \mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta})\varphi_{G,\mathbf{k}}(\mathbf{x}_t)\right|^2.$$

Thus far, it is not clear how the number of the required Fourier coefficients, $\mathcal{K}_{\mathrm{T}}$, should be determined. For convenience, we suggest to choose $\mathcal{K}_{\mathrm{T}}$ when there is not much change between the estimates with $\mathcal{K}_{\mathrm{T}}$ and $\mathcal{K}_{\mathrm{T}} + 1$. That is, given a tolerance level $q\%$, $\mathcal{K}_{\mathrm{T}}$ is chosen when

$$\frac{\|\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}} + 1) - \widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})\|}{\|\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})\|} < q\%,$$

where $\|\cdot\|$ denotes a vector norm.

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 begin our analysis with the case that $m = 1$; the univariate $\mathbf{X}$ is denoted as X (no boldface). The asymptotic properties for the case with multivariate $\mathbf{X}$ are discussed in Section 3.3.

3.1. Consistency

We impose the following conditions.

[A1] The observed data $(\mathbf{y}_t', x_t)'$, $t = 1, \ldots, T$, are independent realizations of $(\mathbf{Y}', X)'$.
[A2] For each $\boldsymbol{\theta} \in \Theta$, $\mathbf{h}(\cdot, \boldsymbol{\theta})$ is measurable, and for each $\mathbf{y} \in \mathbb{R}^r$, $\mathbf{h}(\mathbf{y}, \cdot)$ 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 $\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta})|\mathbf{X}] = \mathbf{0}$.
[A3] $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} |\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta})|^2] < \infty$.
[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 and its derivatives, yet it still permits quite general G functions.

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

$$\varphi_{G,\mathbf{k}}(X) = \int_{[-\pi,\pi]^2} G(A(\mathbf{X}, \boldsymbol{\tau})) \exp(-\mathrm{i}\mathbf{k}'\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}. \tag{13}$$

Here, $\mathbf{k} = (k_0, k_1)'$. 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 $\varphi_{G,\mathbf{k}}(X)$.

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

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

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

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

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

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

where $\xrightarrow{\mathbb{P}}$ *stands for convergence in probability.*

Lemma 3.2 implies

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T} |\mathbf{m}_{G,\mathbf{k},T}(\boldsymbol{\theta})|^2 \xrightarrow{\mathbb{P}} \sum_{k_0,k_1=-\infty}^{\infty} |C_{G,\mathbf{k}}(\boldsymbol{\theta})|^2, \tag{14}$$

uniformly for all $\boldsymbol{\theta}$ in Θ. As $\boldsymbol{\theta}_o$ is the unique minimizer of the right-hand side of (14), the consistency result below follows from Theorem 2.1 of Newey and McFadden (1994).

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

For the estimator $\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_T)$ in (12), note that $\exp(X\tau)$ satisfies [A4] with τ a scalar. It is easy to deduce that Lemma 3.1 holds with $|\varphi_{\exp,k}(X)| \leq \Delta/k$. In analogy with Lemma 3.2, we also have

$$\sum_{k=-\mathcal{K}_T}^{\mathcal{K}_T} |\mathbf{m}_{\exp,k,T}(\boldsymbol{\theta}) - C_{\exp,k}(\boldsymbol{\theta})|^2 \xrightarrow{\mathbb{P}} 0, \tag{15}$$

when $\mathcal{K}_T = o(T)$. The result below follows from (15) and is analogous to Theorem 3.3.

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

3.2. Asymptotic normality

Recall that the Fourier coefficient $C_{G,\mathbf{k}}(\boldsymbol{\theta})$ can be expressed as

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

which is the integral of the product of two functions in $\boldsymbol{\tau}$, i.e., $\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta})G(A(X, \cdot))]$ and $\exp(-\mathrm{i}\mathbf{k}'\cdot)$. To establish asymptotic normality, we work on $\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta})G(A(X, \cdot))]$ 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 $\mathbf{f}$ and $\mathbf{g}$ is $\langle \mathbf{f}, \mathbf{g}\rangle = \int_{-\pi}^{\pi} \mathbf{f}(\tau)'\bar{\mathbf{g}}(\tau)\mathrm{d}\tau$, and the norm induced by the inner product is $\langle \mathbf{f}, \mathbf{f}\rangle^{1/2}$. A random element $\mathbf{U}$ has mean $\mathbb{E}(\mathbf{U})$ if $\mathbb{E}[\langle \mathbf{U}, \mathbf{g}\rangle] = \langle \mathbb{E}(\mathbf{U}), \mathbf{g}\rangle$ for any $\mathbf{g}$ in $L_2[-\pi, \pi]$. The covariance operator $\mathbb{K}$ associated with $\mathbf{U}$ is, for any $\mathbf{g}$ in $L_2[-\pi, \pi]$, $\mathbb{K}\mathbf{g} = \mathbb{E}[\langle \mathbf{U} - \mathbb{E}(\mathbf{U}), \mathbf{g}\rangle(\mathbf{U} - \mathbb{E}(\mathbf{U}))]$ such that

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

with the kernel $\kappa_{ji}(\tau, s) = \mathbb{E}[(U_j(\tau) - \mathbb{E}U_j(\tau))(U_i(s) - \mathbb{E}U_i(s))]$. $\mathbf{U}$ is said to be Gaussian if for any $\mathbf{g}$ in $L_2[-\pi, \pi]$, $\langle \mathbf{U}, \mathbf{g}\rangle$ has a normal distribution on $\mathbb{R}$ with mean $\langle \mathbb{E}(\mathbf{U}), \mathbf{g}\rangle$ and variance $\langle \mathbb{K}\mathbf{g}, \mathbf{g}\rangle$. 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), $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_T)$ must satisfy the first order condition:

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

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

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

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

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T} \mathrm{Re}\Big(\nabla_{\boldsymbol{\theta}}\mathbf{m}_{G,\mathbf{k},T}(\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_T))'[\overline{\mathbf{m}}_{G,\mathbf{k},T}(\boldsymbol{\theta}_o) + \nabla_{\boldsymbol{\theta}}\overline{\mathbf{m}}_{G,\mathbf{k},T}(\boldsymbol{\theta}_T^{\dagger})(\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_T) - \boldsymbol{\theta}_o)]\Big) = \mathbf{0}. \tag{16}$$

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

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

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

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

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

is symmetric and positive definite.

[A8] $T^{-1/2}\sum_{t=1}^{T}\mathbf{h}(\mathbf{y}_t,\boldsymbol{\theta}_o)G(A(x_t,\cdot)) \xrightarrow{D} \mathbb{Z}$, 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

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

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

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 (17). We directly assume functional convergence in [A8] for convenience; this condition is the same as Assumption 11 in Carrasco and Florens (2000). To ensure such convergence, one may also impose primitive conditions on $\mathbf{h}$, G and the data; see, e.g., Chen and White (1998).

To study the behavior of the normalized estimator via (16), we give two limiting results for the terms on the right-hand side of (16).

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

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

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

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

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

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

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

With Lemmas 3.5 and 3.6, (16) can be expressed as

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T)-\boldsymbol{\theta}_o) = -\mathcal{M}_q^{-1}\left[\sum_{k_0,k_1=-\infty}^{\infty}\nabla_{\boldsymbol{\theta}}C_{G,\mathbf{k}}(\boldsymbol{\theta}_o)'\sqrt{T}\overline{\mathbf{m}}_{G,\mathbf{k},T}(\boldsymbol{\theta}_o)\right] + o_{\mathbb{P}}(1). \tag{17}$$

The functional convergence condition [A8] now ensures that the term in the square bracket on the right-hand side of (17) has a limiting normal distribution, which in turn leads to the asymptotic normality of $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T)$.

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

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

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:*

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

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

$$\sum_{k=-\mathcal{K}_T}^{\mathcal{K}_T}\mathrm{Re}(\nabla_{\boldsymbol{\theta}}\mathbf{m}_{\exp,k,T}(\widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_T))'\nabla_{\boldsymbol{\theta}}\overline{\mathbf{m}}_{\exp,k,T}(\boldsymbol{\theta}_T^{\dagger})) \xrightarrow{\mathbb{P}} \sum_{k=-\infty}^{\infty}\nabla_{\boldsymbol{\theta}}C_{\exp,k}(\boldsymbol{\theta}_o)'\nabla_{\boldsymbol{\theta}}\overline{C}_{\exp,k}(\boldsymbol{\theta}_o), \tag{18}$$

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

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

and

$$\sum_{k=-\mathcal{K}_T}^{\mathcal{K}_T}\mathrm{Re}(\nabla_{\boldsymbol{\theta}}\mathbf{m}_{\exp,k,T}(\widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_T))'\sqrt{T}\,\overline{\mathbf{m}}_{\exp,k,T}(\boldsymbol{\theta}_o)) = \sum_{k=-\infty}^{\infty}\nabla_{\boldsymbol{\theta}}C_{\exp,k}(\boldsymbol{\theta}_o)'\sqrt{T}\,\overline{\mathbf{m}}_{\exp,k,T}(\boldsymbol{\theta}_o) + o_{\mathbb{P}}(1). \tag{19}$$

In this case, (17) becomes

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}(\exp,\mathcal{K}_T)-\boldsymbol{\theta}_o) = -\mathcal{M}_q^{-1}\left[\sum_{k=-\infty}^{\infty}\nabla_{\boldsymbol{\theta}}C_{\exp,\mathbf{k}}(\boldsymbol{\theta}_o)'\sqrt{T}\,\overline{\mathbf{m}}_{\exp,\mathbf{k},T}(\boldsymbol{\theta}_o)\right] + o_{\mathbb{P}}(1), \tag{20}$$

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

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

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

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:*

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

For estimation of $\mathcal{V}$ in Corollary 3.8, note from (18) that $\mathcal{M}_q$ can be consistently estimated by

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

From [A8] and (19), $\boldsymbol{\Omega}_q$ can be consistently estimated by the real part of

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

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

3.3. The results for multivariate $\mathbf{X}$

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

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

where $\mathbf{k} = (k_0, k_1, \ldots, k_m)'$. The required conditions for asymptotics are unchanged, except [A4] is changed to [A4′].

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

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

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

Again, let $c(k_i) = |k_i|$ for $k_i \neq 0$ and $c(k_i) = 1$ for $k_i = 0, i = 0, 1, \ldots, m$. Similar to Lemma 3.1, we obtain the following bound on $\varphi_{G,\mathbf{k}}(\mathbf{X})$ when $\mathbf{X}$ is multivariate.

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

With Lemma 3.9, the results below include Theorems 3.3 and 3.7 as special cases. Note that the growth rates of $\mathcal{K}_{\mathrm{T}}$ depend on m, the dimension of $\mathbf{X}$.[4] The results for the specific estimator $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})$ can be obtained similarly.

Theorem 3.10. *Given* [A1]–[A3] *and* [A4′], *if* $\mathcal{K}_{\mathrm{T}} \to \infty$ *and* $\mathcal{K}_{\mathrm{T}} = o(T^{1/(m+1)})$, *then* $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}}) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_o$ *as* $T \to \infty$.

Theorem 3.11. *Given* [A1]–[A3], [A4′] *and* [A5]–[A8], *if* $\mathcal{K}_{\mathrm{T}} \to \infty$ *and* $\mathcal{K}_{\mathrm{T}} = o(T^{1/(2m+2)})$, *then*

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

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:*

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

4. Efficient estimation

It now remains to show how an efficient estimator can be computed; this is the topic to which we now turn. Following Newey (1990, 1993) and Domínguez and Lobato (2004), an efficient estimate may be obtained from the proposed consistent estimate via an additional Newton–Raphson step. That is, an efficient estimator can be computed as

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

where $Q_T(\boldsymbol{\theta})$ is an objective function for the efficient estimator that can locally identify $\boldsymbol{\theta}_o$, and $\nabla_{\boldsymbol{\theta}}Q_T(\boldsymbol{\theta})$ and $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}Q_T(\boldsymbol{\theta})$ are its gradient vector and Hessian matrix, both evaluated at the consistent estimate $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})$. In practice, identifying such an objective function and estimating its gradient and Hessian matrix may not be as straightforward as one would like (Newey, 1990, 1993).

Carrasco and Florens (2000) consider efficient estimation based on the objective function that takes into account the covariance structure:

$$\boldsymbol{\theta}_o = \arg\min_{\boldsymbol{\theta}\in\Theta} \int_{\mathcal{T}} \mathbb{K}^{-1/2}|\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}) \exp(\tau X)]|^2\mathrm{d}P(\tau),$$

where $\mathbb{K}$ is the covariance operator introduced in Section 3.2, and the corresponding estimation method is based on projection along preliminary estimates of the eigenfunctions of $\mathbb{K}$. There are some drawbacks of this approach. First, this estimator depends on various user-chosen parameters and hence is arbitrary to some extent. Second, the generalized inverse of the covariance operator exists only for a subset of Hilbert space, namely, the reproducing kernel Hilbert space. Moreover, it is difficult to generalize their results to allow for multivariate X.

Alternatively, an efficient estimator is readily computed via the conventional two-step GMM method. As $\varphi_{G,\mathbf{k}}(\mathbf{X})$ is complex, we now consider $\varphi^r_{G,\mathbf{k}}(\mathbf{X})$ and $\varphi^i_{G,\mathbf{k}}(\mathbf{X})$, the real and imaginary parts of $\varphi_{G,\mathbf{k}}(\mathbf{X})$.[5] Equivalent to (9), $\boldsymbol{\theta}_o$ can also be identified as

$$\boldsymbol{\theta}_o = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{\mathbf{k}\in\mathcal{S}} |\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta})\varphi^r_{G,\mathbf{k}}(\mathbf{X})]|^2 + |\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta})\varphi^i_{G,\mathbf{k}}(\mathbf{X})]|^2,$$

where the minimum of this objective function is zero. A new set of unconditional moment restrictions now consists of $\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)\varphi^r_{G,\mathbf{k}}(\mathbf{X})] = \mathbf{0}$ and $\mathbb{E}[\mathbf{h}(\mathbf{Y}, \boldsymbol{\theta}_o)\varphi^i_{G,\mathbf{k}}(\mathbf{X})] = \mathbf{0}$ with $\mathbf{k} \in \mathcal{S}$. Given $\varphi^r_{G,\mathbf{k}}(\mathbf{X}) = \varphi^r_{G,-\mathbf{k}}(\mathbf{X})$ and $\varphi^i_{G,\mathbf{k}}(\mathbf{X}) = -\varphi^i_{G,-\mathbf{k}}(\mathbf{X})$ for any $\mathbf{k} \in \mathcal{S}$, some of these unconditional moment restrictions are redundant and can be omitted.

Let $\mathbf{q}_t(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}}) = \mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta}) \otimes \mathbf{Z}_{G,\mathcal{K}_{\mathrm{T}}}(\mathbf{x}_t)$, where $\mathbf{Z}_{G,\mathcal{K}_{\mathrm{T}}}(\mathbf{x}_t)$ is the $(2\mathcal{K}_{\mathrm{T}}+1) \times (4\mathcal{K}_{\mathrm{T}}+1)^m$-dimensional vector that contains $\varphi^r_{G,\mathbf{k}}(\mathbf{x}_t)$ and $\varphi^i_{G,\mathbf{k}}(\mathbf{x}_t)$, where $\mathbf{k} = [k_0, k_1, \ldots, k_m]'$ with $k_0 = 0, 1, \ldots, \mathcal{K}_{\mathrm{T}}$, and $k_i = 0, \pm 1, \ldots, \pm\mathcal{K}_{\mathrm{T}}$ for $i = 1, 2, \ldots, m$. The sample counterpart of the asymptotic covariance matrix of $q_t(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}})$ is

$$\mathbf{V}_T(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{q}_t(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}})\mathbf{q}_t(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}})'.$$

Evaluating the inverse of $\mathbf{V}_T$ at the consistent estimate $\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}})$ and taking the resulting matrix as the weighting matrix, an efficient GMM estimator of $\boldsymbol{\theta}_o$ is

$$\widehat{\boldsymbol{\theta}}^e(G, \mathcal{K}_{\mathrm{T}}) = \arg\min_{\boldsymbol{\theta}\in\Theta} \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{q}_t(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}})\right)' \mathbf{V}_T^{-1}(\widehat{\boldsymbol{\theta}}(G, \mathcal{K}_{\mathrm{T}}), G, \mathcal{K}_{\mathrm{T}}) \times \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{q}_t(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}})\right). \quad (21)$$

In the homoskedasticity case that $\mathbb{E}[\mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta}_o)\mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta}_o)'|\mathbf{X}]$ is constant, $\mathbf{V}_T$ simplifies to

$$\mathbf{V}_T(\boldsymbol{\theta}, G, \mathcal{K}_{\mathrm{T}}) = \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta})\mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta})'\right] \otimes \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{Z}_{G,\mathcal{K}_{\mathrm{T}}}(\mathbf{x}_t)\mathbf{Z}_{G,\mathcal{K}_{\mathrm{T}}}(\mathbf{x}_t)'\right].$$

[4] The dimension m affects the growth rates of $\mathcal{K}_{\mathrm{T}}$ only through the implication rule and the generalized Chebyshev inequality in the proofs.

[5] For example, when X is univariate and G is the exponential function,

$$\varphi^r_{\exp,k}(X) = (-1)^k \frac{2X}{X^2 + k^2} \sinh(\pi X),$$

$$\varphi^i_{\exp,k}(X) = (-1)^k \frac{2k}{X^2 + k^2} \sinh(\pi X),$$

are the real and imaginary parts of $\varphi_{\exp,k}(X)$.

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

Sample	Estimator	$X \sim \mathcal{N}(0,1)$			$X \sim \mathcal{N}(1,1)$		
T		Bias	SE	MSE	Bias	SE	MSE
50	$\hat{\theta}_{\text{NLS}}$	−0.0011	0.0498	0.0025	−0.0002	0.0225	0.0005
	$\hat{\theta}_{\text{OPIV}}$	−0.0168	0.1961	0.0387	−2.2072	0.7923	5.4992
	$\hat{\theta}_{\text{DL}}$	−0.0058	0.1373	0.0189	−0.0011	0.0484	0.0023
	$\hat{\theta}(\exp, \mathcal{K}_{\text{T}})$	−0.0024	0.0776	0.0060	−0.0004	0.0253	0.0006
	$\widehat{\theta}^e(\exp, \mathcal{K}_{\text{T}})$	−0.0017	0.0614	0.0038	−0.0001	0.0248	0.0006
100	$\hat{\theta}_{\text{NLS}}$	−0.0009	0.0347	0.0012	−0.0002	0.0158	0.0002
	$\hat{\theta}_{\text{OPIV}}$	−0.0037	0.0833	0.0070	−2.2266	0.7890	5.5801
	$\hat{\theta}_{\text{DL}}$	0.0021	0.0840	0.0071	−0.0003	0.0336	0.0011
	$\hat{\theta}(\exp, \mathcal{K}_{\text{T}})$	−0.0008	0.0501	0.0025	−0.0001	0.0174	0.0003
	$\widehat{\theta}^e(\exp, \mathcal{K}_{\text{T}})$	−0.0010	0.0394	0.0016	−0.0001	0.0166	0.0003
200	$\hat{\theta}_{\text{NLS}}$	−0.0002	0.0239	0.0006	0.0001	0.0110	0.0001
	$\hat{\theta}_{\text{OPIV}}$	−0.0018	0.0601	0.0036	−2.2382	0.7769	5.6129
	$\hat{\theta}_{\text{DL}}$	−0.0010	0.0580	0.0034	0.0004	0.0244	0.0006
	$\hat{\theta}(\exp, \mathcal{K}_{\text{T}})$	−0.0005	0.0343	0.0012	0.0002	0.0123	0.0002
	$\widehat{\theta}^e(\exp, \mathcal{K}_{\text{T}})$	−0.0003	0.0261	0.0007	0.0001	0.0115	0.0001

NOTE:
1. The DGP is $Y = \theta_o^2 X + \theta_o X^2 + \epsilon$, $\epsilon \sim \mathcal{N}(0, 1)$, where $\theta_o = 1.25$, $X \sim \mathcal{N}(0, 1)$ or $X \sim \mathcal{N}(1, 1)$.
2. We set $\mathcal{K}_{\text{T}} = 5$ for $\widehat{\theta}(\exp, \mathcal{K}_{\text{T}})$ and $\widehat{\theta}^e(\exp, \mathcal{K}_{\text{T}})$.

The inverse of $\mathbf{V}_T$ is easy to compute in practice because the first term in $\mathbf{V}_T$ is positive definite and the inverse of the second term can be obtained by any generalized inverse method. Under the imposed conditions, it is reasonable to expect $\widehat{\boldsymbol{\theta}}^e(G, \mathcal{K}_{\text{T}})$ being as efficient as that of Carrasco and Florens (2000) asymptotically.

By treating $\mathbf{Z}_{G,\mathcal{K}_{\text{T}}}(\mathbf{x}_t)$ as a class of approximating functions, the results in Donald et al. (2003) may be employed to establish the asymptotic properties of the efficient estimator (21).[6] It should be emphasized that, with the proposed unconditional moments, the two-step GMM estimation method is not the only way to obtain an efficient estimator. Other methods, such as empirical likelihood estimation (e.g., Qin and Lawless, 1994) and continuously updated estimation (e.g., Hansen et al., 1996) will also do.

5. Simulations

In this section, we focus on the finite-sample performance of the proposed consistent and efficient estimators: $\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\text{T}})$ and $\widehat{\boldsymbol{\theta}}^e(\exp, \mathcal{K}_{\text{T}})$. We compare their performance with the nonlinear least squares (NLS) estimator:

$$\widehat{\boldsymbol{\theta}}_{\text{NLS}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \frac{1}{T}\sum_{t=1}^{T} |\mathbf{h}(\mathbf{y}_t, \boldsymbol{\theta})|^2,$$

and the DL estimator of Domínguez and Lobato (2004), $\widehat{\boldsymbol{\theta}}_{\text{DL}}$ in (5). 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. In all experiments, the samples are $T = 50, 100, 200$; the number of replications is 5000. In each replication, we randomly draw 10 initial values for all estimators, and for each estimator, the estimate that leads to the smallest value of the objective function is chosen. The data x_t are transformed using a logistic mapping: $\exp(x_t)/[1 + \exp(x_t)]$, so that they are bounded between 0 and 1. Note that we set $\mathcal{K}_{\text{T}} = 5$ for the proposed estimators; the effect of different $\mathcal{K}_{\text{T}}$ on the proposed estimator will be examined in Section 5.4.

5.1. The experiments in Domínguez and Lobato (2004)

Following Domínguez and Lobato (2004), we postulate the following nonlinear model with exogenous regressors:

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

where $\theta_o = 1.25$ 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 $X \sim \mathcal{N}(1, 1)$.

When $X \sim \mathcal{N}(0, 1)$, $\theta_o = 1.25$ 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 $X \sim \mathcal{N}(1, 1)$, in addition to $\theta_o = 1.25$, $\theta = -1.25$ and $\theta = -3$ also satisfy the unconditional moment restriction with the "feasible" optimal instrument. In this case, 1.25 is the global minimum of the NLS objective function; the other two solutions are only local minima. For comparison, our simulations here also includes the optimal instrument variable (OPIV) estimator:

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

which is different from the NLS estimator; cf. Domínguez and Lobato (2004, p. 1608).

The simulation results are summarized in Table 1. In most cases, the NLS estimator outperforms the other estimators in terms of bias, SE and MSE, as it ought to be. On the other hand, $\widehat{\theta}_{\text{OPIV}}$ has severe bias and large SE and is dominated by the other estimators. Note that when $X \sim \mathcal{N}(1, 1)$, the existence of 3 possible solutions (1.25, −1.25 and −3) suggests that the bias of the OPIV estimator should be close to −2.25. This is confirmed in our simulation.[7] It is also clear that the proposed consistent and efficient estimators, $\widehat{\theta}(\exp, \mathcal{K}_{\text{T}})$ and $\widehat{\theta}^e(\exp, \mathcal{K}_{\text{T}})$, both dominate the DL estimator in terms of bias, SE and MSE in all cases. Also, there is no significant difference between these two estimators and the DL estimator in terms of the speed of convergence. When the sample size is not too small ($T = 100, 200$), the performance of the proposed efficient estimator is comparable with that of the NLS estimator. It

[6] Some stronger conditions are needed. For example, when G is the exponential function and X is univariate, Theorems 5.3 and 5.4 in Donald et al. (2003) require the growth rate of $\mathcal{K}_{\text{T}}$ being $o(T^{1/2})$. This is more restrictive than the rate required for the consistent estimator: $\widehat{\boldsymbol{\theta}}(\exp, \mathcal{K}_{\text{T}})$; cf. Corollary 3.4.

[7] Note that Domínguez and Lobato (2004) report a much smaller bias (about −0.4) under the same simulation design.

Table 2
Models with an endogenous regressor.

ρ	Est.	$T = 50$			$T = 100$			$T = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
0.01	$\hat{\theta}_{\mathrm{NLS}}$	0.0014	0.0315	0.0010	0.0010	0.0212	0.0005	0.0015	0.0146	0.0002
	$\hat{\theta}_{\mathrm{DL}}$	−0.0108	0.1129	0.0129	−0.0037	0.0800	0.0064	−0.0006	0.0560	0.0031
	$\hat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0001	0.0557	0.0031	0.0007	0.0368	0.0014	0.0007	0.0239	0.0006
	$\widehat{\theta^e}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0005	0.0492	0.0024	0.0001	0.0331	0.0011	0.0005	0.0218	0.0005
0.1	$\hat{\theta}_{\mathrm{NLS}}$	0.0098	0.0314	0.0011	0.0102	0.0212	0.0006	0.0103	0.0147	0.0003
	$\hat{\theta}_{\mathrm{DL}}$	−0.0137	0.1170	0.0139	−0.0054	0.0826	0.0068	−0.0021	0.0565	0.0032
	$\hat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0025	0.0568	0.0032	−0.0006	0.0358	0.0013	−0.0001	0.0241	0.0006
	$\widehat{\theta^e}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0019	0.0503	0.0025	−0.0001	0.0325	0.0011	0.0003	0.0217	0.0005
0.3	$\hat{\theta}_{\mathrm{NLS}}$	0.0324	0.0314	0.0020	0.0311	0.0210	0.0014	0.0314	0.0145	0.0012
	$\hat{\theta}_{\mathrm{DL}}$	−0.0100	0.1183	0.0141	−0.0067	0.0836	0.0070	−0.0046	0.0574	0.0033
	$\hat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0029	0.0571	0.0033	−0.0023	0.0369	0.0014	−0.0011	0.0242	0.0006
	$\widehat{\theta^e}(\exp, \mathcal{K}_{\mathrm{T}})$	0.0016	0.0512	0.0026	0.0004	0.0335	0.0011	0.0005	0.0220	0.0005
0.5	$\hat{\theta}_{\mathrm{NLS}}$	0.0539	0.0310	0.0039	0.0527	0.0206	0.0032	0.0522	0.0144	0.0029
	$\hat{\theta}_{\mathrm{DL}}$	−0.0120	0.1237	0.0154	−0.0088	0.0852	0.0073	−0.0029	0.0593	0.0035
	$\hat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0049	0.0575	0.0033	−0.0033	0.0360	0.0013	−0.0009	0.0249	0.0006
	$\widehat{\theta^e}(\exp, \mathcal{K}_{\mathrm{T}})$	0.0024	0.0522	0.0027	0.0012	0.0324	0.0011	0.0012	0.0224	0.0005
0.7	$\hat{\theta}_{\mathrm{NLS}}$	0.0746	0.0291	0.0064	0.0736	0.0202	0.0058	0.0731	0.0139	0.0055
	$\hat{\theta}_{\mathrm{DL}}$	−0.0168	0.1268	0.0164	−0.0079	0.0851	0.0073	−0.0051	0.0588	0.0035
	$\hat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0102	0.0606	0.0038	−0.0039	0.0369	0.0014	−0.0026	0.0252	0.0006
	$\widehat{\theta^e}(\exp, \mathcal{K}_{\mathrm{T}})$	0.0011	0.0536	0.0029	0.0020	0.0340	0.0012	0.0007	0.0226	0.0005
0.9	$\hat{\theta}_{\mathrm{NLS}}$	0.0968	0.0287	0.0102	0.0951	0.0187	0.0094	0.0939	0.0131	0.0090
	$\hat{\theta}_{\mathrm{DL}}$	−0.0153	0.1287	0.0168	−0.0082	0.0866	0.0076	−0.0034	0.0587	0.0035
	$\hat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.0112	0.0623	0.0040	−0.0048	0.0372	0.0014	−0.0010	0.0239	0.0006
	$\widehat{\theta^e}(\exp, \mathcal{K}_{\mathrm{T}})$	0.0021	0.0574	0.0033	0.0028	0.0336	0.0011	0.0030	0.0216	0.0005

NOTE:
1. The DGP 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 $\theta_o = 1.25$, $\rho = 0.01, 0.1, 0.3, 0.5, 0.7, 0.9$, and $X \sim \mathcal{N}(0, 1)$ is independent of ϵ and ν.
2. We set $\mathcal{K}_{\mathrm{T}} = 5$ for $\widehat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$ and $\widehat{\theta^e}(\exp, \mathcal{K}_{\mathrm{T}})$.

is somewhat surprising to see that, compared with the proposed consistent estimator, our efficient estimator has not only smaller SE and MSE but also slightly smaller bias in many cases. The trade-off between SE and bias can be seen in the experiments in Section 5.3.

5.2. 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 $\theta_o = 1.25, \rho = 0.01, 0.1, 0.3, 0.5, 0.7, 0.9$, 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.

It is clear that the bias of these estimators all increases with ρ. In particular, the NLS estimator has very large biases, and such biases do not diminish when the sample size increases. This should not be surprising because the NLS estimator is inconsistent (due to the endogenous regressor). On the other hand, the proposed consistent and efficient estimators perform remarkably well. They have much smaller bias than the NLS estimator, and they again outperform the DL estimator in terms of bias, SE, and MSE for any ρ and any sample size. Although the NLS estimator typically has a smaller SE, the proposed estimators may yield smaller MSE when the correlation between ϵ and ν is not too small (e.g., $\rho \geq 0.3$).

5.3. 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 $\theta_o = 1.25$, X is the uniform random variable on $(-1, 1)$ and independent of ϵ, and σ^2 =1, 4 and 9. It can be verified that there are 3 solutions to the unconditional moment restriction resulted from the "feasible" optimal instrument $(2\theta X + X^2)$: $\theta = 1.25$ and $(-25 \pm \sqrt{145})/40$, where 1.25 is the global minimum.

The results are summarized in Table 3. We note first that, 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. The proposed consistent estimator has smaller biases than all other estimators in all cases, except its bias is slightly larger than the NLS estimator when $\sigma^2 = 1$ and $T = 200$. In terms of MSE, the proposed consistent estimator dominates the DL estimator in all cases and outperforms the OPIV estimator when σ^2 is not too large. Although the proposed efficient estimator has larger bias than $\widehat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$ in most cases, it still outperforms the other estimators in terms of bias in all cases (except when $T = 50$ and $\sigma^2 = 1$). Moreover, the efficient estimator has the smallest MSE in all cases with $T = 100, 200$, and its MSE is only slightly larger than the NLS estimator for $T = 50$. As far as MSE is concerned, the proposed efficient estimator is to be preferred to the other estimators.

5.4. 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

Table 3
Models with different disturbance variances.

σ^2	Estimator	$T = 50$			$T = 100$			$T = 200$		
		Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
1	$\widehat{\theta}_{\mathrm{NLS}}$	−0.4144	0.9391	1.0535	−0.2227	0.7166	0.5631	−0.0680	0.4055	0.1690
	$\widehat{\theta}_{\mathrm{OPIV}}$	−1.0993	0.9027	2.0230	−1.1143	0.9086	2.0669	−1.1370	0.9230	2.1444
	$\widehat{\theta}_{\mathrm{DL}}$	−0.8468	1.0712	1.8643	−0.6325	0.9782	1.3567	−0.4303	0.8550	0.9160
	$\widehat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.3146	1.0875	1.2814	−0.1543	0.8050	0.6717	−0.0685	0.4931	0.2478
	$\widehat{\theta}^e(\exp, \mathcal{K}_{\mathrm{T}})$	−0.4241	0.9645	1.1100	−0.2105	0.7089	0.5467	−0.0559	0.3707	0.1405
4	$\widehat{\theta}_{\mathrm{NLS}}$	−0.7696	1.1993	2.0303	−0.6318	1.1061	1.6223	−0.4155	0.9399	1.0559
	$\widehat{\theta}_{\mathrm{OPIV}}$	−1.1114	0.9403	2.1192	−1.0835	0.9315	2.0415	−1.1021	0.9157	2.0529
	$\widehat{\theta}_{\mathrm{DL}}$	−1.2274	1.1969	2.9387	−1.0644	1.1360	2.4232	−0.8628	1.0637	1.8757
	$\widehat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.5048	1.5702	2.7199	−0.3924	1.2976	1.8374	−0.2346	1.0388	1.1340
	$\widehat{\theta}^e(\exp, \mathcal{K}_{\mathrm{T}})$	−0.7614	1.2632	2.1750	−0.5727	1.0948	1.5262	−0.3573	0.8932	0.9252
9	$\widehat{\theta}_{\mathrm{NLS}}$	−0.9124	1.3041	2.5328	−0.8078	1.2110	2.1187	−0.6299	1.1001	1.6067
	$\widehat{\theta}_{\mathrm{OPIV}}$	−1.1412	0.9744	2.2517	−1.1136	0.9458	2.1345	−1.0870	0.9238	2.0347
	$\widehat{\theta}_{\mathrm{DL}}$	−1.3675	1.3043	3.5709	−1.2492	1.1965	2.9918	−1.0814	1.1259	2.4369
	$\widehat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$	−0.5511	1.8035	3.5556	−0.4478	1.5662	2.6530	−0.3624	1.3195	1.8720
	$\widehat{\theta}^e(\exp, \mathcal{K}_{\mathrm{T}})$	−0.8602	1.3805	2.6453	−0.7479	1.2245	2.0584	−0.5786	1.0817	1.5046

NOTE:
1. The DGP is $Y = \theta_o^2 X + \theta_o X^2 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$, where $\theta_o = 1.25$, X is the uniform random variable on $(-1, 1)$ and independent of ϵ, and $\sigma^2 = 1, 4$ and 9.
2. We set $\mathcal{K}_{\mathrm{T}} = 5$ for $\widehat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$ and $\widehat{\theta}^e(\exp, \mathcal{K}_{\mathrm{T}})$.

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

$\mathcal{K}_{\mathrm{T}}$	Bias	Bias (+%)	SE	SE (+%)	MSE	MSE (+%)
	$T = 100$					
1	−0.0033	–	0.0378	–	0.0014	–
2	−0.0032	−0.6704	0.0373	−1.3585	0.0014	−2.6884
3	−0.0032	−0.2632	0.0371	−0.5527	0.0014	−1.0981
4	−0.0032	−0.1372	0.0370	−0.2945	0.0014	−0.5858
5	−0.0032	−0.0835	0.0369	−0.1817	0.0014	−0.3617
6	−0.0032	−0.0560	0.0369	−0.1230	0.0014	−0.2447
7	−0.0032	−0.0401	0.0369	−0.0886	0.0014	−0.1763
8	−0.0032	−0.0301	0.0368	−0.0668	0.0014	−0.1330
9	−0.0032	−0.0234	0.0368	−0.0521	0.0014	−0.1038
10	−0.0032	−0.0187	0.0368	−0.0418	0.0014	−0.0832
15	−0.0032	−0.0560	0.0368	−0.1258	0.0014	−0.2504
20	−0.0032	−0.0279	0.0367	−0.0631	0.0014	−0.1256
$\hat{\theta}_{\mathrm{DL}}$	−0.0081		0.0868		0.0076	
	$T = 200$					
1	−0.0011	–	0.0251	–	0.0006	–
2	−0.0011	0.8887	0.0248	−1.3657	0.0006	−2.7036
3	−0.0011	0.3617	0.0247	−0.5580	0.0006	−1.1091
4	−0.0011	0.1945	0.0246	−0.2979	0.0006	−0.5928
5	−0.0011	0.1210	0.0245	−0.1840	0.0006	−0.3663
6	−0.0011	0.0824	0.0245	−0.1245	0.0006	−0.2480
7	−0.0011	0.0597	0.0245	−0.0897	0.0006	−0.1788
8	−0.0012	0.0452	0.0245	−0.0677	0.0006	−0.1348
9	−0.0012	0.0354	0.0245	−0.0528	0.0006	−0.1053
10	−0.0012	0.0285	0.0244	−0.0424	0.0006	−0.0844
15	−0.0012	0.0863	0.0244	−0.1276	0.0006	−0.2540
20	−0.0012	0.0436	0.0244	−0.0640	0.0006	−0.1275
$\hat{\theta}_{\mathrm{DL}}$	−0.0025		0.0577		0.0033	

NOTE:
1. The DGP is the same as that for Table 2.
2. +% stands for the percentage change when $\mathcal{K}_{\mathrm{T}}$ increases.
3. The bias (SE, MSE) look the same across different $\mathcal{K}_{\mathrm{T}}$ because the program recorded only 4 digits after the decimal points. These numbers are actually different, as can be seen in the column of percentage changes.

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

$\mathcal{K}_{\mathrm{T}}$	Bias	Bias (+%)	SE	SE (+%)	MSE	MSE (+%)
	$T = 100$					
1	−0.0047	–	0.0379	–	0.0015	–
2	−0.0047	−0.4598	0.0374	−1.2201	0.0014	−2.4021
3	−0.0047	−0.1770	0.0372	−0.4955	0.0014	−0.9786
4	−0.0047	−0.0909	0.0371	−0.2638	0.0014	−0.5214
5	−0.0047	−0.0547	0.0371	−0.1627	0.0014	−0.3216
6	−0.0047	−0.0363	0.0370	−0.1100	0.0014	−0.2175
7	−0.0047	−0.0258	0.0370	−0.0792	0.0014	−0.1567
8	−0.0047	−0.0193	0.0370	−0.0597	0.0014	−0.1181
9	−0.0047	−0.0149	0.0370	−0.0466	0.0014	−0.0922
10	−0.0047	−0.0119	0.0370	−0.0374	0.0014	−0.0739
15	−0.0047	−0.0354	0.0369	−0.1124	0.0014	−0.2223
20	−0.0047	−0.0175	0.0369	−0.0564	0.0014	−0.1115
$\hat{\theta}_{\mathrm{DL}}$	−0.0089	–	0.0860	–	0.0075	–
	$T = 200$					
1	−0.0021	–	0.0257	–	0.0007	–
2	−0.0021	−1.0072	0.0253	−1.3333	0.0006	−2.6445
3	−0.0021	−0.4112	0.0252	−0.5452	0.0006	−1.0856
4	−0.0021	−0.2183	0.0251	−0.2911	0.0006	−0.5804
5	−0.0021	−0.1342	0.0251	−0.1798	0.0006	−0.3587
6	−0.0021	−0.0905	0.0250	−0.1217	0.0006	−0.2429
7	−0.0021	−0.0650	0.0250	−0.0877	0.0006	−0.1751
8	−0.0021	−0.0489	0.0250	−0.0662	0.0006	−0.1321
9	−0.0020	−0.0381	0.0250	−0.0517	0.0006	−0.1031
10	−0.0020	−0.0305	0.0250	−0.0414	0.0006	−0.0827
15	−0.0020	−0.0916	0.0250	−0.1247	0.0006	−0.2489
20	−0.0020	−0.0458	0.0249	−0.0626	0.0006	−0.1249
$\hat{\theta}_{\mathrm{DL}}$	−0.0048		0.0592		0.0035	

NOTE:
1. The DGP is the same as that in Table 2.
2. +% stands for the percentage change when $\mathcal{K}_{\mathrm{T}}$ increases.
3. The bias (SE, MSE) look the same across different $\mathcal{K}_{\mathrm{T}}$ because the program recorded only 4 digits after the decimal points. These numbers are actually different, as can be seen in the column of percentage changes.

Section 5.2, 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 DL estimator and $\widehat{\theta}(\exp, \mathcal{K}_{\mathrm{T}})$ with $\mathcal{K}_{\mathrm{T}} = 1, 2, \ldots, 10, 15, 20$. We do not consider the NLS estimator because its performance is too poor when regressor is endogenous. To ease our computation, we do not simulate the efficient estimator here. 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 0.46%, SE decreases 1.22%, and MSE decreases 2.40% 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 small; in most cases, such changes are less than 0.1% when $\mathcal{K}_T$ 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 $\mathcal{K}_T$ can only result in marginal improvements on the bias and SE. Note that the proposed estimator again dominates the DL estimator in terms of bias, SE and MSE in all cases.

6. Concluding remarks

This paper is concerned with consistent and efficient estimation of conditional moment restrictions without assuming the parameters can be identified by the implied unconditional moments. We propose an approach to constructing unconditional moments that can identify the parameter of interest. The consistent and efficient estimators are then readily computed using the conventional GMM method. Our simulations confirm that the proposed estimators perform very well in finite samples and compare favorably with existing estimators, such as that proposed by Domínguez and Lobato (2004). It must be emphasized that we do not have to confine ourselves with GMM estimation. Based on the proposed moment conditions, other estimation methods, such as the empirical likelihood method, can also be employed to obtain consistent and/or efficient estimators.

The proposed estimator may be further improved. First, it is important to establish a criterion determining the (optimal) number of the required Fourier coefficients, $\mathcal{K}_T$, in the objective function. Second, as different GCR functions result in different sets of unconditional moment conditions and hence different estimators, it would be very useful if we know, in practice, how to choose a better one among such estimators. Moreover, it is interesting to examine if the proposed method remains effective when the instruments are "weak". Chao and Swanson (2005) show that consistent estimation may be possible when the number of weak instruments increases to infinity at some suitable rate. As our approach provides a systematic way to increase the number of unconditional moments so as to identify parameters, it may still work even when instruments are weak. These topics all require thorough analysis and hence are left to future research.

Acknowledgments

We would like to thank 3 anonymous referees for their constructive comments. We are indebted to Zongwu Cai, John Chao, Jin-Chuan Duan, James Hamilton, Jinyong Hahn, Yongmiao Hong, Ying-Ying Lee, Joon Park, M. Hashem Pesaran, Peter C. B. Phillips, Max Stinchcombe, Allan Timmermann, Yoon-Jae Whang, and Hal White for their comments on early drafts of this paper. We also thank the seminar participants at Boston College, Rice, Toronto, UCSD, U. of Hong Kong, USC, UT Austin, and the SETA 2007 meeting in Hong Kong for their inputs. 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.

Appendix

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

$$\begin{aligned}
&\varphi_{G,\mathbf{k}}(X)\\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} G(\tau_0+\tau_1 X)\exp(-\mathrm{i}k_0\tau_0)\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_0\mathrm{d}\tau_1\\
&= \int_{-\pi}^{\pi}\left[\int_{-\pi}^{\pi} G(\tau_0+\tau_1 X)\exp(-\mathrm{i}k_0\tau_0)\mathrm{d}\tau_0\right]\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1.
\end{aligned}$$

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

$$\begin{aligned}
&\int_{-\pi}^{\pi} G(\tau_0+\tau_1 X)\exp(-\mathrm{i}k_0\tau_0)\mathrm{d}\tau_0\\
&= \frac{\mathrm{i}}{k_0}\Bigg\{\underbrace{(-1)^{k_0}[G(\pi+\tau_1X)-G(-\pi+\tau_1X)]}_{Q_1(\boldsymbol{\tau})}\\
&\quad \underbrace{-\int_{-\pi}^{\pi} G_0(\tau_0+\tau_1X)\exp(-\mathrm{i}k_0\tau_0)\mathrm{d}\tau_0}_{Q_2(\boldsymbol{\tau})}\Bigg\}.
\end{aligned}$$

Then,

$$\varphi_{G,\mathbf{k}}(X) = \frac{\mathrm{i}}{k_0}\int_{-\pi}^{\pi}[Q_1(\boldsymbol{\tau})-Q_2(\boldsymbol{\tau})]\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1,$$

so that

$$\begin{aligned}
|\varphi_{G,\mathbf{k}}(X)| \le \frac{1}{|k_0|}\Bigg\{&\left|\int_{-\pi}^{\pi} Q_1(\boldsymbol{\tau})\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\right|\\
&+\left|\int_{-\pi}^{\pi} Q_2(\boldsymbol{\tau})\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\right|\Bigg\}.
\end{aligned}$$

Again by integration by parts,

$$\begin{aligned}
&\int_{-\pi}^{\pi} Q_1(\boldsymbol{\tau})\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\\
&= \frac{(-1)^{k_0}\mathrm{i}}{k_1}\Bigg\{(-1)^{k_1}[G(\pi+\pi X)-G(-\pi+\pi X)\\
&\quad - G(\pi-\pi X)+G(-\pi-\pi X)]\\
&\quad - \int_{-\pi}^{\pi}\left[G_1(\pi+\tau_1X)-G_1(-\pi+\tau_1X)\right]\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\Bigg\},
\end{aligned}$$

and

$$\begin{aligned}
&\int_{-\pi}^{\pi} Q_2(\boldsymbol{\tau})\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\\
&= \frac{\mathrm{i}}{k_1}\Bigg\{(-1)^{k_1}\int_{-\pi}^{\pi}[G_0(\tau_0+\pi X)\\
&\quad - G_0(\tau_0-\pi X)]\exp(-\mathrm{i}k_0\tau_0)\mathrm{d}\tau_0\\
&\quad - \int_{-\pi}^{\pi}\left(\int_{\pi}^{\pi} G_{01}(\tau_0+\tau_1X)\exp(-\mathrm{i}k_0\tau_0)\mathrm{d}\tau_0\right)\\
&\quad \times \exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\Bigg\}.
\end{aligned}$$

Given [A4], we have

$$\begin{aligned}
&\left|\int_{-\pi}^{\pi} Q_1(\boldsymbol{\tau})\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\right|\\
&\le \frac{1}{|k_1|}\Bigg[4\sup_{\boldsymbol{\tau}\in\mathcal{T}}|G(\tau_0+\tau_1X)|
\end{aligned}$$

[8] In the lower panel of Table 4, the bias actually increases with $\mathcal{K}_T$. This ill behavior may be due to the convergence criterion in our procedure; the criterion for the gradient of estimated coefficients is set to 10^{-4}.

$$+2\int_{-\pi}^{\pi}\sup_{\tau\in\mathcal{T}}|G_1(\tau_0+\tau_1 X)||\exp(-\mathrm{i}k_1\tau_1)|\mathrm{d}\tau_1\Bigg]$$

$$\le \frac{\Delta}{|k_1|},$$

and

$$\left|\int_{-\pi}^{\pi} Q_2(\boldsymbol{\tau})\exp(-\mathrm{i}k_1\tau_1)\mathrm{d}\tau_1\right|$$

$$\le \frac{1}{|k_1|}\left[2\int_{-\pi}^{\pi}\sup_{\tau\in\mathcal{T}}|G_0(\tau_0+\tau_1 X)||\exp(-\mathrm{i}k_0\tau_0)|\mathrm{d}\tau_0\right.$$

$$+\int_{-\pi}^{\pi}\left(\int_{\pi}^{\pi}\sup_{\tau\in\mathcal{T}}|G_{01}(\tau_0+\tau_1 X)||\exp(-\mathrm{i}k_0\tau_0)|\mathrm{d}\tau_0\right)$$

$$\left.\times|\exp(-\mathrm{i}k_1\tau_1)|\mathrm{d}\tau_1\right]$$

$$\le \frac{\Delta}{|k_1|}.$$

It follows that $|\varphi_{G,\mathbf{k}}(X)| \le \Delta/(|k_0||k_1|)$ for $k_0, k_1 \ne 0$. Similarly, we can show that $|\varphi_{G,\mathbf{k}}(X)| \le \Delta/|k_1|$ for $k_0 = 0$ and $k_1 \ne 0$ and that $|\varphi_{G,\mathbf{k}}(X)| \le \Delta/|k_0|$ for $k_0 \ne 0$ and $k_1 = 0$. Also, it is clear that $|\varphi_{G,0}(X)| \le \Delta$. The proof is thus complete. □

Proof of Lemma 3.2. Again let Δ denote a generic constant whose value varies in different cases. Define

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

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

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

Under [A1], these bounds lead to

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

$$\le \frac{4\Delta}{T}\sum_{k_0=1}^{\mathcal{K}_T}\frac{1}{k_0^2}\sum_{k_1=1}^{\mathcal{K}_T}\frac{1}{k_1^2}+\frac{2\Delta}{T}\sum_{k_0=1}^{\mathcal{K}_T}\frac{1}{k_0^2}+\frac{2\Delta}{T}\sum_{k_1=1}^{\mathcal{K}_T}\frac{1}{k_1^2}+\frac{\Delta}{T}$$

$$\le \frac{\Delta}{T},$$

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

$$\mathbb{P}\left[\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_{G,\mathbf{k},t}\right|^2 \ge \varepsilon\right]$$

$$\le \sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}\mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\eta}_{G,\mathbf{k},t}\right|^2 \ge \frac{\varepsilon}{(2\mathcal{K}_T+1)^2}\right]$$

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

$$\le \frac{(2\mathcal{K}_T+1)^2}{\varepsilon}\frac{\Delta}{T},$$

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

Proof of Theorem 3.3. The proposed estimator, $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T)$, is the solution to the left-hand side of (14). Hence, it must converge to the unique minimizer, $\boldsymbol{\theta}_o$, of the right-hand side of (14) by Theorem 2.1 of Newey and McFadden (1994). □

Proof of Corollary 3.4. Given $G(A(X,\boldsymbol{\tau})) = \exp(X\tau)$, we have from the text that (15) holds when $\mathcal{K}_T = o(T)$. Analogous to (14), we obtain

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

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 $\mathcal{K}_T = o(T^{1/4})$, $\widehat{\boldsymbol{\theta}}(G,\mathcal{K}_T) \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_o$. Hence, $\boldsymbol{\theta}_T^{\dagger} \to \boldsymbol{\theta}_o$. With [A6], we can apply a standard argument to get

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

$$\nabla_{\boldsymbol{\theta}}\overline{\mathbf{m}}_{G,\mathbf{k},T}(\boldsymbol{\theta}_T^{\dagger}) - \nabla_{\boldsymbol{\theta}}\overline{\mathbf{m}}_{G,\mathbf{k},T}(\boldsymbol{\theta}_o) \xrightarrow{\mathbb{P}} \mathbf{0}.$$

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

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}\nabla_{\boldsymbol{\theta}}C_{G,\mathbf{k}}(\boldsymbol{\theta}_o)'\nabla_{\boldsymbol{\theta}}\overline{C}_{G,\mathbf{k}}(\boldsymbol{\theta}_o)$$

$$\to \sum_{k_0,k_1=-\infty}^{\infty}\nabla_{\boldsymbol{\theta}}C_{G,\mathbf{k}}(\boldsymbol{\theta}_o)'\nabla_{\boldsymbol{\theta}}\overline{C}_{G,\mathbf{k}}(\boldsymbol{\theta}_o).$$

Therefore, it suffices to show that

$$\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}(\nabla_{\boldsymbol{\theta}}\mathbf{m}_{G,\mathbf{k},T}(\boldsymbol{\theta}_o)'\nabla_{\boldsymbol{\theta}}\overline{\mathbf{m}}_{G,\mathbf{k},T}(\boldsymbol{\theta}_o)$$

$$-\nabla_{\boldsymbol{\theta}}C_{G,\mathbf{k}}(\boldsymbol{\theta}_o)'\nabla_{\boldsymbol{\theta}}\overline{C}_{G,\mathbf{k}}(\boldsymbol{\theta}_o)) \xrightarrow{\mathbb{P}} \mathbf{0}.$$

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

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

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

$$\mathbb{P}\left\{\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}|\eta_{i,\mathbf{k}}'\nabla_{\theta_j}\overline{\mathbf{m}}_{\mathbf{k},T} + \nabla_{\theta_i}C_{\mathbf{k}}'\bar{\eta}_{j,\mathbf{k}}| \ge \epsilon\right\}$$

$$\le \sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}\mathbb{P}\left\{|\eta_{i,\mathbf{k}}'\nabla_{\theta_j}\overline{\mathbf{m}}_{\mathbf{k},T} + \nabla_{\theta_i}C_{\mathbf{k}}'\bar{\eta}_{j,\mathbf{k}}| \ge \frac{\epsilon}{(2\mathcal{K}_T+1)^2}\right\}$$

$$\le \frac{(2\mathcal{K}_T+1)^2}{\epsilon}\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}\mathbb{E}[|\eta_{i,\mathbf{k}}'\nabla_{\theta_j}\overline{\mathbf{m}}_{\mathbf{k},T} + \nabla_{\theta_i}C_{\mathbf{k}}'\bar{\eta}_{j,\mathbf{k}}|]$$

$$\le \frac{(2\mathcal{K}_T+1)^2}{\epsilon}\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T}[\mathbb{E}|\eta_{i,\mathbf{k}}|^2]^{1/2}[\mathbb{E}|\nabla_{\theta_j}\overline{\mathbf{m}}_{\mathbf{k},T}|^2]^{1/2}$$

$$+[\mathbb{E}|\nabla_{\theta_i}C_{\mathbf{k}}|^2]^{1/2}[\mathbb{E}|\bar{\eta}_{j,\mathbf{k}}|^2]^{1/2}.$$

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

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

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

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

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

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

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

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

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

since

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

and

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

where, by invoking the multiplication theorem,

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

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

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

where the last inequality, given [A1], is due to the fact that

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

From the proof of Lemma 3.5 we have

$$\mathbb{E}|\eta_{i,\mathbf{k}}|^2 \le \frac{\Delta}{Tc(k_0)^2c(k_1)^2},$$

and

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

It follows that

$$\mathbb{P}\left\{\sum_{k_0,k_1=-\mathcal{K}_T}^{\mathcal{K}_T} |\eta_{i,\mathbf{k}}'\sqrt{T}\overline{\mathbf{m}}_{G,\mathbf{k},T}(\boldsymbol{\theta}_o)| \ge \epsilon\right\} \le \frac{(2\mathcal{K}_T+1)^2}{\epsilon}\frac{\Delta}{\sqrt{T}}.$$

The proof is complete because this bound can be made arbitrarily small when $\mathcal{K}_T = o(T^{1/4})$ and $T \to \infty$. □

Proof of Theorem 3.7. From [A8], we know that $T^{-1/2}\sum_{t=1}^{T}\mathbf{h}(\mathbf{y}_t,\boldsymbol{\theta}_o)G(A(x_t,\cdot)) \xrightarrow{D} \mathbb{Z}$, 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

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

The assertion follows from (17). □

Proof of Corollary 3.8. In this case, [A8] ensures that $T^{-1/2}\sum_{t=1}^{T}\mathbf{h}(\mathbf{y}_t,\boldsymbol{\theta}_o)\exp(x_t,\cdot) \xrightarrow{D} \mathbb{Z}$, 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 (20). □

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71, 1795–1843.

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. Cambridge University Press, New York.

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

Chao, J.C., Swanson, N.R., 2005. Consistent estimation with large number of weak instruments. Econometrica 73, 1673–1692.

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

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

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

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

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

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

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

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

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

Hansen, B.E., West, K.D., 2002. Generalized method of moments and macroeconomics. Journal of Business and Economic Statistics 20, 460–469.

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

Kitamura, Y., Tripathi, G., Ahn, H., 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: Maddala, G., Rao, C., Vinod, H. (Eds.), Handbook of Statistics, vol. 11. Elsevier Science, pp. 419–454.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis test. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. IV. Elsevier Science, pp. 2111–2245.

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

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

Stuart, R.D., 1961. An Introduction to Fourier Analysis. Halsted Press, New York.

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