Consistent estimation of asset pricing models using generalized spectral estimator

Jinho Choi*
Indiana University

April 10, 2009
Work in progress

Abstract

This paper essentially extends the generalized spectral estimation of Berkowitz (2001) to provide a consistent generalized spectral estimator (GSE), considering all the information available, possibly with infinite dimensions, based upon Escanciano (2006). Our estimator can entertain the strengths of the Berkowitz-GSE over the standard GMM. In contrast, more importantly, the newly proposed estimator has consistency which the Berkowitz-GSE is deficient in, overcoming Domínguez and Lobato's (2004) critique on the identifiability of the GMM approach. Furthermore, our estimator is more general, based upon fairly relaxed assumptions for its asymptotic behaviors, than the Berkowitz-GSE. Finally, as an empirical application, using the proposed estimation strategy, we estimate the standard consumption-based asset pricing model in Hansen and Singleton (1982) to investigate the possibility that the equity premium puzzle may be due to underidentification of risk aversion parameters.

JEL Classification: C13; C22
Keywords: Identification; Conditional moment; Spectral analysis

*Department of Economics, 105 Wylie Hall, 100 S. Woodlawn, Bloomington, IN 47405, U.S.A.; choi25@indiana.edu. I would like to thank my advisor, Professor Juan Carlos Escanciano for many helpful comments and suggestions. All errors are mine.

1 Introduction

Since introduced by Hansen (1982), the generalized method of moments (GMM) has been widely used to estimate conditional moment restrictions implied by economic theories. Among a variety of the advantages, GMM has immediately gained popularity in econometrics mainly because no distributional assumptions are needed. For instance, as micro-founded macroeconomics becomes standard in the decade, a great number of literature in finance and macroeconomics has employed the GMM approach to estimate a specific type of conditional moment restrictions called the Euler equations, which characterize the agents' decision-making resulting from the utility maximization; for illustration, see Hansen and Singleton (1982), Harvey (1991), and Gali and Gertler (1999). Although most applications of GMM are based upon the time domain approach, Berkowitz's (2001) work is remarkable in the sense that he proposes a frequency domain version of GMM, named generalized spectral estimator (GSE).[1] In estimating parameters of interest, both approaches convert conditional moment restrictions into unconditional moments, whereas subsequent procedures would be entirely different across the two approaches. To clarify this point, suppose we derive arbitrary conditional moment restrictions from the theory as follows:

$$E[h(Y_t, \theta_0)|X_t] = 0 \quad a.s. \text{ for a unique value } \theta_0 \in \Theta, \text{ where } \Theta \subset \mathbb{R}^p \tag{1}$$

Then, to apply the standard GMM or GSE appoach, econometricians take into account unconditional moments as their moment conditions:

$$E[h(Y_t, \theta_0)g(X_t)] = 0 \quad a.s. \quad \text{for any given } g(\cdot) \tag{2}$$

Given the population condition (2), the standard estimation strategy in the literature is firmly based upon the assumption that θ_0 is globally identified, arbitrarily selecting a finite number of unconditional moments out of infinite candidates of $g(X_t)$, and then minimizing a sample analogue of the objective function to yield GMM-type estimators. However, Domínguez and Lobato (2004) point out that the key assumption in the GMM literature may be seriously flawed and thus give rises to nontrivial problems in terms of consistency because the unconditional moments utilize fairly limited information on the data generating process,

[1] In his earlier working paper version, Berkowitz (1996) named the proposed methodology as *Spectral GMM*. However, Chacko and Viceira (2003) also use the term to denote their estimation strategy for continuous-time stochastic models based upon the characteristic function. In order to avoid confusion, we do not use the term *Spectral GMM* in this paper.

showing that their minimum distance estimator (DL) outperforms the GMM estimator in identifying parameters across different types of data generating process. Considering this possibility of underidentification may have substantial implications on the empirical literature relying on the GMM-type estimation, providing some clues to solve several interesting problems, including the equity premium puzzle.

This paper basically extends Berkowitz (2001) to propose a generalized spectral estimator, possibly with a infinite dimension, based upon Escanciano (2006), which employs a generalized spectral distribution to provide goodness-of-fit tests for the parametric conditional mean. Our estimator can entertain the strengths of the Berkowitz' GSE over the traditional GMM (*e.g.*, focusing on a subset of frequencies, no need to consider the weighting matrix). In contrast, from the perspective of Domínguez and Lobato (2004), the proposed estimator is consistent whereas the Berkowitz's GSE may not be consistent as a result of lack of identification. In this sense, our estimator can be considered as *spectral-DL*. As a simple application, we estimate the classical consumption-based asset pricing model in Hansen and Singleton (1982) to compare our estimation strategy with the two existing methods: the Berkowitz's GSE, and the standard GMM.

The paper proceeds as follows. Section 2 overviews identification issues in the GMM-type estimation, illustrating with an example in Domínguez and Lobato (2004). Section 3 proposes an alternative generalized spectral estimator to Berkowitz's (2001) GSE and provides the asymptotic theory. Section 4 presents the estimation and testing results for a consumption-based asset pricing model. Section 5 concludes.

2 Identification issues in GMM

With the rational expectation prevailing in several fields of economic theory, conditional moment restrictions are widely being used to describe model equilibrium, in which researchers are eventually interested. In the literature, the most popular estimation strategy for conditional moment restrictions has been generalized method of moments (GMM) proposed by Hansen (1982). However, despite several advantages, to implement GMM in practice, finding appropriate instruments with relevance and validity is fairly challenging. In this regard, there is a growing number of literature raising a variety of questions about the identifiability of GMM. For instance, a line of literature actively examines on namely 'weak identification problem', caused by faint relevances between instruments and endogenous variables. See Stock and Wright (2000), Stock et al. (2002), Andrews and Stock (2005).

Furthermore, more recent work suspects that the GMM approach may even fail to identify parameters especially when a model is defined by conditional moment conditions. Domínguez and Lobato (2004) show that the GMM's key assumption of global identification may be seriously flawed, providing a consistent estimator. In addition, Hsu and Kuan (2008) use Fourier-coefficient based to propose a consistent estimator, which is favorably compared with Domínguez and Lobato's.

In this section we use Domínguez and Lobato's (2004) simple illustration to explore potential identification failures of the GMM-type estimation. Consider a univariate random variable Y with the conditional mean of $E(Y|X) = \theta_0^2 X + \theta_0 X^2$. Assume the true value of θ_0 equals 5/4 and $V(Y|X)$ is constant. Furthermore, suppose that an econometrician correctly specifies the model and chooses the optimal instrument $W = 2\theta X + X^2$. Then, he can construct the unconditional moment condition:

$$\begin{aligned} E[(Y - \theta^2 X - \theta X^2)W] &= E[(Y - \theta^2 X - \theta X^2)(2\theta X + X^2)] \\ &= E[(E[Y|X] - \theta^2 X - \theta X^2)(2\theta X + X^2)] \\ &= E[(\theta_0 - \theta)\{X^4 + (\theta_0 + 3\theta)X^3 + 2\theta(\theta_0 + \theta)X^2\}] \\ &= 0 \end{aligned} \tag{3}$$

Then, when the conditioning variable X follows an $N(0,1)$, the last equality in the condition (3) holds only if $\theta = \theta_0 = +5/4$, which implies identification. In contrast, when X follows an $N(1,1)$, either $\theta = -3$ or $-5/4$ as well as the true value of $+5/4$ makes the unconditional moment equal to zero, showing no identification or underidentification. This simple example manifests the case that the global identification assumption in GMM may not hold:

$$E[h(Y_t, \theta)g(X_t)] = 0 \quad a.s. \text{ for some } g(X) \quad \nRightarrow \quad \theta = \theta_0 \tag{4}$$

Intuitively, we can interpret the case (4) against the identification assumption in GMM as follows. For any given conditional moment restriction (1), one can generate an infinite number of unconditional moment restrictions (or instruments $g(X)$) in (2). However, in practice, selecting only a few instruments may lead to inconsistent estimation because replacing conditional moments by unconditional moments may require losing crucial information from the original restrictions. To overcome the risk of potential underidentification, Domínguez and Lobato also propose an alternative estimator using the whole information about θ_0 in the conditional moment restriction (1). Using Theorem 16.10 (iii) in Billingsley (1995), one can obtain the following equivalence:

$$E[h(Y_t,\theta_0)|X_t] = 0 \ \ a.s. \ \Leftrightarrow H(\theta_0,x) = 0 \ \text{ for almost all } \ x \in \mathbb{R}^d \tag{5}$$

where $H(\theta,x) = E[h(Y_t,\theta)I(X_t \le x)]$ and $I(\cdot)$ indicator function. Then, the population parameter θ_0 can be recovered by minimizing the measure of the distance of $H(\theta,x)$ from 0,

$$\theta_0 = \arg\min_{\theta\in\Theta} \int H(\theta,x)^2 dP_{X_t}(x) \tag{6}$$

where P_{X_t} is the probability density function of the random vector X_t.

Therefore, corresponding to (6), Domínguez and Lobato propose a minimum distance estimator (DL),

$$\hat{\theta}_{DL} = \arg\min_{\theta\in\Theta} \frac{1}{n^3} \sum_{l=1}^{n} \Big(\sum_{t=1}^{n} h(Y_t,\theta)I(X_t \le X_l) \Big)^2, \tag{7}$$

which is consistent and asymptotically normal. Furthermore, using a simulation study (Table I, p. 1608), they show that in terms of bias, standard error and mean square error, the DL estimator outperforms the GMM estimator for either $X \sim N(0,1)$ or $X \sim N(1,1)$, which we analytically considered above. With this background, the following section discusses spectral estimators in the frequency domain framework, corresponding to the standard GMM.

3 Generalized spectral estimation

3.1 Generalized spectral estimators (GSE)

Given that the time domain framework is dominant in econometric analysis, why should we still need to pay attention to the frequency domain approach as considered in this paper? It is because some difficult problems under one framework may be easily resolved using the other. Furthermore, in general, using the frequency domain allows us to assess the contributions of individual frequencies to overall identification, as well as to reduce computational burden relative to the time domain approach.

Motivated by Durlauf (1991), Berkowitz (2001) proposes a generalized spectral estimator in the frequency domain, corresponding to standard GMM estimators in the time domain approach. Under his framework, selecting lags of the Euler residual, $h(Y_{t-j},\theta)$ as instruments replaces the conditional moment restriction (1) with

$$E[h(Y_t,\theta_0)h(Y_{t-j},\theta_0)] = 0 \quad a.s. \text{ for } \ j \geq 1, \tag{8}$$

implying that the autocovariance function $\gamma(j) = 0$ for $j \geq 1$ and thus making the associated spectral density

$$f_{h_t(\theta_0)}(u) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma(j)e^{-iju} = \frac{\gamma_{\theta_0}(0)}{2\pi} \equiv \frac{\sigma^2}{2\pi}, \tag{9}$$

where u denotes frequency and $h_t(\theta) \equiv h(Y_t,\theta)$. From (9), we can observe that at $\theta = \theta_0$, the spectral density of $h(Y_t,\theta_0)$ is flat, i.e., $f_{h_t(\theta_0)}(u) = \sigma^2/2\pi$ over its entire support, otherwise deviating from the constant value. Using this fact, one can formulate the distance of a spectral density from the constant as follows:

$$S(\lambda) = \int_0^{\lambda\pi}\Big(\frac{f_{h_t}(u)}{\sigma^2} - \frac{1}{\pi}\Big)du, \quad \lambda \in (0,1) \tag{10}$$

Then, using the usual Cramér-von Mises (CvM) norm to measure the distance, Berkowitz proposes a generalized spectral density estimator (BGSE).

$$\widehat{\theta}_{BGSE} = \arg\min_{\theta\in\Theta}\int_0^1 \widehat{S}(\lambda)^2 d\lambda \ \text{ where } \widehat{S}(\lambda) = \int_0^{\lambda\pi}\Big(\frac{\widehat{f}_{h_t}(u)}{\widehat{\sigma}^2} - \frac{1}{\pi}\Big)du \tag{11}$$

From the viewpoint of identification, however, BGSE cannot avoid the Domínguez and Lobato's critique because the condition (8) still assumes that the autocovariance function yields zero if and only if $\theta = \theta_0$. In what follows, we will show that this identification assumption is not always valid, thus highlighting the potential absence of identification in the context of Berkowitz (2001).

Let us consider a simple linear process.

$$Y_t = \theta_0 X_t + \varepsilon_t \ \text{ with } E[\varepsilon_t|\mathfrak{F}_{t-1}] = 0, \tag{12}$$

where $\mathfrak{F}_{t-1}$ is the σ-field generated by the conditioning set $I_{t-1} = (X_{t-1}, Y_{t-1}, X_{t-2}, Y_{t-2}, ...)'$. Then, assuming the model is correctly specified, we obtain $E[Y_t|\mathfrak{F}_{t-1}] = \theta_0 X_t$, and $\varepsilon_t(\theta) = Y_t - \theta_0 X_t$. Given the condition, let us check if the Berkowitz's identification assumption (8) is valid, *i.e.*

$$E[\varepsilon_t(\theta)\varepsilon_{t-j}(\theta)] = 0 \ \ a.s. \text{ for } \ j \geq 1 \Longleftrightarrow \theta = \theta_0 \tag{13}$$

Hence, we can rewrite the autocovariance function as follows:

$$\begin{aligned} E[\varepsilon_t(\theta)\varepsilon_{t-j}(\theta)] &= E[(Y_t-\theta X_t)(Y_{t-j}-\theta X_{t-j})] \\ &= E[(E(Y_t|\mathfrak{F}_{t-1})-\theta X_t)(Y_{t-j}-\theta X_{t-j})] \\ &= (\theta_0-\theta)^2 E[X_t X_{t-j}] + (\theta_0-\theta)E[X_t\varepsilon_{t-j}] \end{aligned} \tag{14}$$

Then, identification may fail because the equality in (14) holds when for all $j \geq 1$,

$$\theta = \theta_0 \quad \text{or} \quad \theta = \theta_0 + \frac{E[X_t\varepsilon_{t-j}]}{E[X_t X_{t-j}]} \qquad \text{if } E[X_t X_{t-j}] \neq 0 \tag{15}$$

Specifically, for an AR(1) process ($i.e.$,$X_t = Y_{t-1}$), we can easily show that the autocovariance function (14) equals to zero when $\theta = 1/\theta_0$ as well as the true θ_0.[2] Therefore, we can rewrite (13) as

$$E[\varepsilon_t(\theta)\varepsilon_{t-j}(\theta)] = 0 \quad a.s. \text{ for } \ j \geq 1 \Longleftrightarrow \theta = \theta_0 \quad \text{or} \quad 1/\theta_0$$

which implies that identification may fail.

In order to verify the possibility of underidentification, we simulate AR(1) process $y_t = 0.8y_{t-1}+\varepsilon_t,\ \varepsilon_t \sim N(0,1)$. To minimize the dependence upon the selection of initial values y_0, we generate N_1 observations and then wash out the initial N_0. Figure 1 presents the objective function of squared sample autocovariance function $E_n[\varepsilon_t(\theta)\varepsilon_{t-j}(\theta)]$ along the grid of possible values for $\theta \in [0.5, 1.5]$, setting with $N_1 = 2000$, $N_0 = 1000$, $j = 5$. As the figure shows, we can verify that the objective function is minimized at zero when $\theta = 1.25$ $(= 1/0.8)$, as well as the true value of 0.8. Accordingly, to exclude the possibility that parameters of interest may not be identified, we extend Berkowitz (2001) to propose a consistent estimator considering all the information available, possibly with infinite dimensions, based upon the

[2] The two terms associated with the second solution in (15) can be obtained as follows:

$$\begin{aligned} E[X_t\varepsilon_{t-j}] &= E[Y_{t-1}\varepsilon_{t-j}] = E[(\sum_{h=0}^{\infty}\theta_0^h\varepsilon_{t-h-1})\varepsilon_{t-j}] \\ &= \theta_0^{j-1}\sigma^2, \qquad \text{for} \quad |\theta_0|<1 \end{aligned} \tag{16}$$

$$\begin{aligned} E[X_t X_{t-j}] &= E[Y_{t-1}Y_{t-j-1}] = E[(\sum_{h=0}^{\infty}\theta_0^h\varepsilon_{t-h-1})(\sum_{k=0}^{\infty}\theta_0^k\varepsilon_{t-k-j-1})] \\ &= \sum_{h=0}^{\infty}\sum_{k=0}^{\infty}\theta_0^h\theta_0^k Cov(\varepsilon_{t-h-1},\varepsilon_{t-k-j-1}) = \Big(\frac{\theta_0^j}{1-\theta_0^2}\Big)\sigma^2 \end{aligned} \tag{17}$$

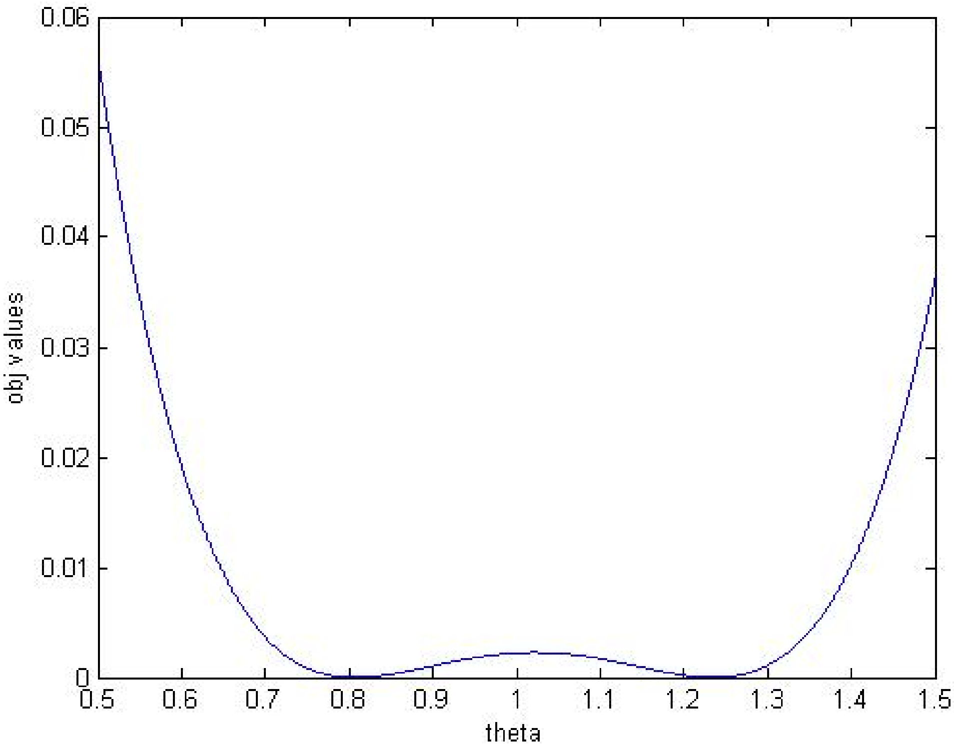

Figure 1: Underidentification in AR(1) with $\theta_0 = 0.8$

testing methodology by Escanciano (2006).

Escanciano (2006) introduces the use of a generalized spectral distribution for testing martingale difference hypothesis (1). Among several advantages, using the spectral distribution allows us to skip the choice of any kernel and bandwidth for testing. Moreover, unlike Berkowitz (2001), we can escape the potential identification problem if converting the test statistics proposed by Escanciano (2006) into minimization criteria.

To obtain a consistent estimator, we follow the notations and procedure to derive the integrated generalized spectral tests in Escanciano (2006). Let $\{(Y_t, \mathbf{X}'_{t-1})\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic time series process defined on the probability space $(\Omega, \mathcal{F}, P)$, where $Y_t \in \mathbb{R}$ dependent variable and $\mathbf{Z}_{t-1} = (Y_{t-1}, \mathbf{X}'_{t-1})' \in \mathbb{R}^m$, $m \in \mathbb{N}$, is the explanatory random vector including the lags of Y_t and $\mathbf{X}_t$. Furthermore, we denote the conditioning set at time $t-1$ as $\mathbf{I}_{t-1} = (\mathbf{Z}'_{t-1}, \mathbf{Z}'_{t-2}, ...)'$. Then, let us consider a parameterized conditional moment restriction implied from an economic theory:

$$E[h(Y_t, \theta_0)|\mathbf{I}_{t-1}] = 0 \;\; a.s. \;\; \text{for a unique } \theta_0 \in \Theta \subset \mathbb{R}^p, \tag{18}$$

which is equivalent to

$$E[h(Y_t, \theta_0)|\mathbf{Z}_{t-j}] = 0 \;\; a.s. \; \forall \text{ j} \geq 1, \text{ for a unique } \theta_0 \in \Theta \subset \mathbb{R}^p. \tag{19}$$

Then, by selecting an appropriate function from the family of functions $\mathcal{F} = \{w(\mathbf{Z}, \mathbf{x}) : \mathbf{x} \in \Upsilon \subset \mathbb{R}^s\}$ satisfying Lemma 1 in Escanciano (2006), we can rewrite the restriction (19) using a generalized measure of dependence $\gamma_{j,w}(\cdot)$ as

$$\gamma_{j,w}(\mathbf{x}, \theta_0) = E[h(Y_t, \theta_0)w(\mathbf{Z}_{t-j}, \mathbf{x})] = 0 \;\; a.e. \text{ in } \Upsilon \subset \mathbb{R}^s, \; s \in \mathbb{N}, \text{ j} \geq 1. \tag{20}$$

While among popular examples of $w(\mathbf{Z}, \mathbf{x})$ are the exponential functions or indicator functions, we maintain the general notation in this derivation. For a list of the literature using different weighting functions, see Escanciano and Velasco (2006).

Then, applying the Fourier transform to the functions $\left\{\gamma_{j,w}(\mathbf{x}, \theta)\right\}_{j=-\infty}^{\infty}$, we obtain a spectral density

$$f_w(u, \mathbf{x}, \theta) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_{j,w}(\mathbf{x}, \theta) e^{-iju}, \quad \forall u \in [-\pi, \pi], \; \mathbf{x} \in \Upsilon$$

where $i = \sqrt{-1}$, and u denotes frequency.

Following Escanciano (2006), we construct a generalized spectral distribution function as

$$\begin{aligned} H_w(\lambda, \mathbf{x}, \theta) &= 2\int_0^{\lambda\pi} f_w(u, \mathbf{x}, \theta) du, \quad \lambda \in (0, 1) \\ &= \gamma_{0,w}(\mathbf{x}, \theta)\lambda + 2\sum_{j=1}^{\infty} \gamma_{j,w}(\mathbf{x}, \theta) \frac{\sin j\pi\lambda}{j\pi} \end{aligned} \tag{21}$$

As with (9), a careful investigation reveals that evaluated at $\theta = \theta_0$, the spectral distribution function yields a constant value over the entire support:

$$H_w(\lambda, \mathbf{x}, \theta_0) = \gamma_{0,w}(\mathbf{x}, \theta_0)\lambda \tag{22}$$

Then, let us consider a sample $\{Y_{t-1}, \widehat{\mathbf{I}}_{t-1}\}_{t=1}^{n}$ where $\widehat{\mathbf{I}}_{t-1} = (\mathbf{Z}'_{t-1}, \mathbf{Z}'_{t-2}, ..., \mathbf{Z}'_0)'$, and denote the sample conditional moment restriction as $h_t(\theta) \equiv \widehat{h}_t(Y_t, \theta)$. Then the sample analogue of (21) becomes

$$\widehat{H}_w(\lambda, \mathbf{x}, \theta) = \widehat{\gamma}_{0,w}(\mathbf{x}, \theta)\lambda + 2\sum_{j=1}^{n} \widehat{\gamma}_{j,w}(\mathbf{x}, \theta)(n_j/n)^{1/2}\frac{\sin j\pi\lambda}{j\pi} \tag{23}$$

where $\widehat{\gamma}_{j,w}(\mathbf{x}, \theta) = \frac{1}{n_j}\sum_{t=j}^{n} h_t(\theta)w(\mathbf{Z}_{t-j}, \mathbf{x})$ for $j \geq 1$, $n_j = n - j + 1$, and $(n_j/n)^{1/2}$ a finite-sample correction factor.

In the spirit of Berkowitz (2001), we can formulate the deviation of the sample spectral distribution $\widehat{H}_w(\lambda, \mathbf{x}, \theta)/\widehat{\gamma}_{0,w}(\mathbf{x}, \theta)$ from the constant $\lambda = H_w(\lambda, \mathbf{x}, \theta_0)/\gamma_{0,w}(\mathbf{x}, \theta_0)$, or equivalently the distance between $\widehat{H}_w(\lambda, \mathbf{x}, \theta)$ and $\widehat{H}_{0,w}(\lambda, \mathbf{x}, \theta) = \widehat{\gamma}_{0,w}(\mathbf{x}, \theta)\lambda$:

$$\begin{aligned} S_{n,w}(\lambda, \mathbf{x}, \theta) &= \left(\frac{n}{2}\right)^{1/2}\left\{\widehat{H}_w(\lambda, \mathbf{x}, \theta) - \widehat{H}_{0,w}(\lambda, \mathbf{x}, \theta)\right\} \\ &= 2\sum_{j=1}^{n} {n_j}^{1/2}\widehat{\gamma}_{j,w}(\mathbf{x}, \theta)\frac{\sqrt{2}\sin j\pi\lambda}{j\pi} \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} h_t(\theta)q_{t,w}(\lambda, \mathbf{x}, \theta), \end{aligned} \tag{24}$$

where $q_{t,w}(\lambda, \mathbf{x}, \theta) \equiv \sum_{j=1}^{t}(n/n_j)^{1/2}\frac{\sqrt{2}\sin j\pi\lambda}{j\pi}w(\mathbf{Z}_{t-j}, \mathbf{x})$. Therefore, using the Cramér-von Mises (CvM) norm, we can measure the distance $S_{n,w}(\cdot)$ as

$$D^2_{n,w}(\theta) = \int_{\Pi} |S_{n,w}(\lambda, \mathbf{x}, \theta)|^2 \, W(d\mathbf{x})d\lambda \tag{25}$$

where $\Pi = [0, 1] \times \Upsilon$ and $W(\cdot)$ is an integrating function associated with the weight family $\mathcal{F}$ defined above. Given the last expression for $S_{n,w}(\lambda, \mathbf{x}, \theta)$ in (24), the CvM norm can be considered as application of the Integrated Conditional Moment (ICM) statistic proposed by Bierens (1982). Accordingly, it follows from the minimization of the norm (25) that we obtain a generalized spectral estimator (CGSE) as

$$\widehat{\theta}_{CGSE} = \arg\min_{\theta\in\Theta} \; D^2_{n,w}(\theta) \tag{26}$$

Specifically, if we choose the exponential function for $w(\mathbf{Z}_{t-j}, \mathbf{x})$ in the dependence measure (20), (*i.e.*, $w(\mathbf{Z}_{t-j}, \mathbf{x}) = \exp(i\mathbf{x}'\mathbf{Z}_{t-j}),\ \mathbf{x} \in \mathbb{R}^m$) with a selection of the cumulative distribution function of a standard normal random variable for the integrating function in $W(\cdot)$, then

(26) can be rewritten as

$$
\begin{aligned}
\widehat{\theta}_{CGSE} &= \arg\min_{\theta\in\Theta} \ D^2_{n,C}(\theta) \\
&= \arg\min_{\theta\in\Theta} \int_{\Pi} |S_{n,C}(\lambda,\mathbf{x},\theta)|^2 \, d\Phi(\mathbf{x})d\lambda \\
&= \arg\min_{\theta\in\Theta} \sum_{j=1}^{n} \frac{1}{\widehat{\gamma}_{0,C} n_j (j\pi)^2} \sum_{t=j}^{n}\sum_{s=j}^{n} h_t(\theta)h_s(\theta)\exp\{-\frac{1}{2}(\mathbf{Z}_{t-j}-\mathbf{Z}_{s-j})^2\}
\end{aligned}
\tag{27}
$$

Note that given the exponential weighting function, we can easily obtain moment conditions associated with BGSE by differentiating characteristic functions. In this sense, we find that CGSE can be considered as a generalized version of BGSE. Furthermore, the proposed estimator attains useful asymptotic behaviors such as consistency and asymptotic normality as provided in the next subsection.

3.2 Asymptotic theory

3.2.1 Consistency

Assumption 1 *The parametric space Θ is compact in $\mathbb{R}^p$. The true parameter θ_0 belongs to the interior of Θ.*

Assumption 2 $\{Y_t, \mathbf{Z}_{t-1}\}_{t\in\mathbb{Z}}$ *is a strictly stationary and ergodic process.*

Assumption 3 $h(Y_t,\theta)$ *is continuous at each $\theta\in\Theta$ with probability one and satisfies* $E[\sup_{\theta\in\Theta}|h(Y_t,\theta)|]<\infty$.

Assumption 4 $E[h(Y_t,\theta)|X_t]=0$ *a.s. if and only if* $\theta=\theta_0$.

Theorem 1 *Let Assumptions 1-4 hold. Then*

$$\widehat{\theta}_{CGSE} \to_{a.s.} \theta_0.$$

Proof. Due to (22), the population objective function $D^2_C(\theta)$ is uniquely minimized at $\theta_0\in\Theta$. Furthermore, it follows from Assumptions 1-4 and stanardard M-estimator theory that the sample analogue $D^2_{n,C}(\theta)$ converges uniformly in probability to $D^2_C(\theta)$. Then, by Amemiya (1985, Theorem 4.1.1), it completes the proof. ■

3.2.2 Asymptotic normality

Denote $\eta = (\lambda, \mathbf{x}')' \in \Pi$ and consider the process $S_n(\eta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_t(\theta) q_t(\eta)$ where $q_t(\eta) = q_{t,C}(\lambda, \mathbf{x}, \theta)$. Then under standard regularity conditions, similarly to Bierens (1990),

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -D(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_t(\theta_0) \frac{\partial}{\partial \theta'} h_t(\theta_0) + o_p(1)$$

and

$$D(\theta) = E[(\frac{\partial}{\partial \theta'} h_t(\theta))(\frac{\partial}{\partial \theta'} h_t(\theta))'], \quad b(\theta, \eta) = -E[\frac{\partial}{\partial \theta'} h_t(\theta) q_t(\eta)].$$

Hence, by Lemma 3 in Bierens (1990), we can rewrite $S_n(\eta)$ as

$$S_n(\eta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_t(\theta) q_t(\eta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_t(\theta_0) \phi_t(\eta) + o_p(1),$$

where $\phi_t(\eta) = q_t(\eta) + b(\theta_0, \eta)' D(\theta_0)^{-1} \frac{\partial}{\partial \theta'} h_t(\theta_0)$.

Assumption 5 *$h(Y_t, \theta)$ is once continuously differentiable in a neighborhood of θ_0, satisfying $E[\sup_{\theta \in \Xi_0} |\dot{h}(Y_t, \theta)|] < \infty$ where Ξ_0 is a neighborhood of θ_0 and $\dot{h}(Y_t, \theta) = \frac{\partial}{\partial \theta} h(Y_t, \theta)$.*

Assumption 6 *$h(Y_t, \theta_0)$ is a martingale difference sequence with respect to $\{\mathbf{Z}_s, s \leq t\}$.*

Assumption 7 *$E[h^4(Y_t, \theta_0) \|X_t\|^{1+\delta}] < \infty$ and the density of the conditioning variables given the history is continuous and bounded.*

Theorem 2 *Let Assumptions 1-7 hold. Then*

$$\sqrt{n}(\widehat{\theta}_{CGSE} - \theta_0) \rightarrow_d N(0, \Sigma)$$

where

$$\Sigma = \left(\int \dot{H} \dot{H}' d\Phi \right)^{-1} \int \int \dot{H}(\eta_1) \dot{H}'(\eta_2) \Gamma(\eta_1, \eta_2) d\Phi(\eta_1) d\Phi(\eta_2) \left(\int \dot{H} \dot{H}' d\Phi \right)^{-1}$$

with $\dot{H}(\eta) = E[\dot{h}_t(\theta_0) q_t(\eta)]$ and covariance matrix $\Gamma(\eta_1, \eta_2) = p \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h_t^2 \phi_t(\eta_1) \phi_t(\eta_2)$.

Proof.
Let us denote $S_n(\theta) = S_{n,C}(\lambda, \mathbf{x}, \theta)$. Then the minimization of the objective function in (27)

yields the following first order conditions:

$$\int_{\Pi} \dot{S}_n(\widehat{\theta}) S_n(\widehat{\theta}) d\Phi(\mathbf{x}) d\lambda = 0 \tag{28}$$

Then, by the mean value theorem, we obtain

$$\int_{\Pi} \dot{S}_n(\widehat{\theta}) S_n(\theta_0) d\Phi(\mathbf{x}) d\lambda + \int_{\Pi} \dot{S}_n(\widehat{\theta}) \dot{S}_n(\overline{\theta}) d\Phi(\mathbf{x}) d\lambda \times (\widehat{\theta} - \theta_0) = 0$$

where $\overline{\theta} = \rho\theta_0 + (1-\rho)\widehat{\theta}$ for some random $\rho \in [0,1]$. Therefore, we can rewrite the first order conditions as

$$\begin{aligned}
\sqrt{n}(\widehat{\theta} - \theta_0) &= -\left[\int_{\Pi} \dot{S}_n(\widehat{\theta}) \dot{S}_n(\overline{\theta}) d\Phi(\mathbf{x}) d\lambda\right]^{-1} \sqrt{n}\left[\int_{\Pi} \dot{S}_n(\widehat{\theta}) S_n(\theta_0) d\Phi(\mathbf{x}) d\lambda\right] \\
&= -\left[\int_{\Pi} \left\{\frac{1}{n}\sum_{t=1}^{n} \dot{h}_t(\widehat{\theta}) q_t(\eta)\right\}\left\{\frac{1}{n}\sum_{t=1}^{n} \dot{h}_t(\overline{\theta}) q_t(\eta)\right\} d\Phi(\mathbf{x}) d\lambda\right]^{-1} \times \\
&\quad \times \left[\int_{\Pi} \left\{\frac{1}{n}\sum_{t=1}^{n} \dot{h}_t(\widehat{\theta}) q_t(\eta)\right\}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h_t(\theta_0) q_t(\eta)\right\} d\Phi(\mathbf{x}) d\lambda\right].
\end{aligned}$$

Using the continuous mapping theorem, combined with Assumption 5, Lemma 1 and 2 below completes the proof. ■

Lemma 1 *Let θ^* be a consistent estimator of θ_0, and Assumptions 1-7 hold. Then*

$$\frac{1}{n}\sum_{t=1}^{n} \dot{h}_t(\theta^*) q_t(\eta) \rightarrow_{a.s.} \dot{H}(\eta) = E[\dot{h}_t(\theta_0) q_t(\eta)] \quad \textit{uniformly in } \eta.$$

Proof. Let us denote $\dot{H}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} \dot{h}_t(\theta) q_t(\eta)$ and $\dot{H}(\theta) = E[\dot{h}_t(\theta) q_t(\eta)]$. Then consider

$$\begin{aligned}
\left|\dot{H}_n(\theta^*) - \dot{H}(\theta_0)\right| &\leq \left|\dot{H}_n(\theta^*) - \dot{H}(\theta^*)\right| + \left|\dot{H}(\theta^*) - \dot{H}(\theta_0)\right| \\
&\leq \sup_{\theta\in\Theta}\left|\dot{H}_n(\theta) - \dot{H}(\theta)\right| + \left|\dot{H}(\theta^*) - \dot{H}(\theta_0)\right| \rightarrow_{a.s.} 0,
\end{aligned}$$

which is implied by the uniform law of large numbers, the consistency of θ^* and the continuous mapping theorem under Assumptions 1-7. ■

Lemma 2 *Let Assumptions 1-7 hold. Then*

$$S_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_t(\theta_0) q_t(\eta) \Rightarrow S$$

where $\Rightarrow$ *denotes weak convergence in* $C[\Pi]$ *and* S *is a Gaussian process on* Π, *with zero mean and covariance function* $\Gamma(\eta_1, \eta_2)$.

Proof. Let us denote $S_n(\theta)$ as S_n. Then, by the Prohorov's Theorem, it suffices to show that the finite-dimensional distributions *(fidis)* of the random function S_n converges to normal distribution and that S_n is asymptotically tight. The first part of the proof can be easily obtained by applying a version of martingale difference central limit theorem. See Bierens (1994, Theorem 6.1.7). To prove the tightness of S_n, we define $\upsilon_n(\eta)$ as tight random functions on Π such that $P[S_n = \upsilon_n] \geq 1 - \varepsilon$ for an arbitrary ε and need to show that υ_n is tight. For the tightness of υ_n, according to by the Kolmogorov-Cencov criterion, we need to show that there exists a constant C such that

$$E\left|\upsilon_n(\eta_0)\right|^{\gamma} \leq C \quad \text{and} \quad E\left|\upsilon_n(\eta_1) - \upsilon_n(\eta_2)\right|^{\gamma} \leq C \left\|\eta_1 - \eta_2\right\|^{k+\delta}$$

for $\forall \eta_0, \eta_1, \eta_2 \in \Pi$, $\exists \gamma, \delta > 0$,and k is the dimension of Π. Following Bierens and Ploberger (1997), define the stopping time $\tau(M) = \sup\{t \leq n | A_t(\eta_0) \leq nM,\ B_t \leq nM\}$ for an arbitrary $\eta_0 \in \Pi$ and $M > 0$, with $A_t(\eta) = \sum_{j=1}^{t} h_j^2 \phi_j(\eta)^2$, $B_t = \sum_{j=1}^{t} h_j^2 K_j{}^2$. Then, using Burkholder's inequality to υ_n proves the first condition of the criterion when $\gamma = 2k + 2$:

$$E\left|\upsilon_n(\eta_0)\right|^{2k+2} \leq C_{2k+2}(1/n^{k+1}) E(\textstyle\sum_{t=1}^{\tau(M)} h_t^2 \phi_t(\eta_0)^2)^{k+1} \leq C_{2k+2} M^{k+1}.$$

For the proof of the second part, by applying Burkholder's inequality, the Lipschitz condition and the definition of the stopping time $\tau(M)$, we can obtain

$$\begin{aligned}
E\left|\upsilon_n(\eta_1) - \upsilon_n(\eta_2)\right|^{2k+2} &= (1/n^{k+1}) E\left(\left|\textstyle\sum_{t=1}^{\tau(M)} h_t(\phi_t(\eta_1) - \phi_t(\eta_2))\right|^{2k+2}\right) \\
&\leq C_k (1/n^{k+1}) E\left(\textstyle\sum_{t=1}^{\tau(M)} h_t^2 (\phi_t(\eta_1) - \phi_t(\eta_2))^2\right)^{k+1} \\
&\leq C_k (1/n^{k+1}) E\left(\textstyle\sum_{t=1}^{\tau(M)} h_t^2 K_t^2\right)^{k+1} \left\|\eta_1 - \eta_2\right\|^{2k+2} \\
&\leq C_k \left\|\eta_1 - \eta_2\right\|^{2k+2} M^{k+1}.
\end{aligned}$$

which implies that υ_n is tight, and therefore S_n converges weakly to a Gaussian process S. For details, see Bierens and Ploberger (1997) ■

4 Application: consumption-based CAPM (In progress)

It is widely recognized that taking data to standard consumption-based asset pricing models using GMM generates too high risk aversion (*'Equity premium puzzle'*). Among a huge number of alternative models, models with habit persistence (Campbell and Cochrane (1999)) or long-run risk factor (Bansal and Yaron (2004)) have drawn much attentions from researchers. However, Berkowitz (2001) proposes such a high level of risk aversion may be due to the noise in the high frequency data, using the generalized spectral estimation technique. We employ the proposed consistent generalized spectral estimator to estimate a standard consumption-based CAPM, fixing the potential lack of identification in Berkowitz (2001).

[Estimation results will be included.]

5 Concluding remarks

This paper proposes a consistent generalized spectral estimator for models defined by conditional moment restrictions. By employing the frequency domain approach, our estimator is close to a generalized spectral estimator proposed by Berkowitz (2001), but resolves the lack of identification problem, caused by the global identification assumption in the GMM literature. Although Domínguez and Lobato (2004) point out this issue and provide a consistent estimator in the time domain framework, their estimator has a serious drawback: incompatible with high dimensional data. In contrast, our estimator can be applied to the applications with high dimension data. However, our estimator has limitations of using pairwise dependence measures whereas it provides more precise estimation rather than using the whole information set. This is left for future research.

References

Amemiya, T. (1985): *Advanced Econometrics*, Mass: Harvard University Press.

Andrews, D. W. and J. H. Stock (2005): "Inference with Weak Instruments," Cowles Foundation Discussion Papers 1530, Cowles Foundation, Yale University.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Berkowitz, J. (1996): "Generalized spectral estimation," Finance and Economics Discussion Series 96-37, Board of Governors of the Federal Reserve System (U.S.).

——— (2001): "Generalized Spectral Estimation of the Consumption-based Asset Pricing Model," *Journal of Econometrics*, 104, 269–288.

Bierens, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

——— (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–58.

Bierens, H. J. and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1152.

Billingsley, P. (1995): *Probability and Measure*, New York: Wiley and Sons.

Campbell, J. Y. and J. Cochrane (1999): "Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

Chacko, G. and L. M. Viceira (2003): "Spectral GMM estimation of continuous-time processes," *Journal of Econometrics*, 116, 259–292.

Domínguez, M. A. and I. N. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72, 1601–1615.

Durlauf, S. N. (1991): "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics*, 50, 355–376.

Escanciano, J. C. (2006): "Goodness-of-Fit Tests for Linear and Nonlinear Time Series Models," *Journal of the American Statistical Association*, 101, 531–541.

Escanciano, J. C. and C. Velasco (2006): "Generalized Spectral Tests for the Martingale Difference Hypothesis," *Journal of Econometrics*, 134, 151–185.

Gali, J. and M. Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 44, 195–222.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–54.

Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–86.

Harvey, C. R. (1991): "The World Price of Covariance Risk," *Journal of Finance*, 46, 111–57.

Hsu, S.-H. and C.-M. Kuan (2008): "Estimation of Conditional Moment Restrictions without Assuming Parameter Identifiability in the Implied Unconditional Moments," *manuscript*.

Stock, J. H. and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20, 518–29.