2. GMM Estimation Techniques and Properties

2.1. Large Sample Theory and Inference in GMM Estimation

2.1.1. Parameter Estimation using GMM

We follow the framework proposed by Martínez-Iriarte et al. (2020) who consider the econometric estimation using the GMM methodology, as briefly described below. Let $\theta \in \Theta$ denote a $p-$dimensional vector of parameters partitioned into $\theta = (\vartheta', \psi')'$ of dimensions of p_ϑ and p_ψ, respectively.

Denote with

$$F_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f_t(\theta), \tag{2.1}$$

to represent the sample moments, where $f_t(\theta)$ is a $k-$dimensional vector-valued function of data and parameters with $k \geq p$ and $\mathbb{E}\big[f_t(\theta)\big] = 0$ as the true value of θ. Moreover, we let with $r(\theta)$ to be a known function of the parameters such that $r : \Theta \to \mathbb{R}^q, q \leq p_\psi$.

Suppose that $f_t(\vartheta, .)$ and $r(\vartheta, .)$ are continuously differentiable with respect to ψ, and let

$$J_T(\theta) = \frac{\partial F_T(\theta)}{\partial \psi'} \quad \text{and} \quad R(\theta) = \frac{\partial r(\theta)}{\partial \psi'} \tag{2.2}$$

Moreover, we denote with $\hat{V}_f(\theta)$ the $(k \times k)$ matrix that is positive definite almost surely, and define the GMM objective function as below

$$S_T(\vartheta, \psi) = F_T(\vartheta, \psi)' \hat{V}_f(\vartheta, \tilde{\psi})^{-1} F_T(\vartheta, \psi), \tag{2.3}$$

Furthermore, suppose that the constrained GMM estimator of ψ given ϑ exists and is given by the following expression

$$\hat{\psi}(\theta) = \underset{\psi}{\operatorname{argmin}}\, F_T(\vartheta, \psi)' \hat{V}_f(\vartheta, \tilde{\psi})^{-1} F_T(\vartheta, \psi). \tag{2.4}$$

We also simplify the notation as below

$$\hat{\psi} \equiv \hat{\psi}(\vartheta), \quad \hat{r}(\vartheta) = r(\vartheta, \hat{\psi}) \tag{2.5}$$

In addition we consider $\hat{C}(\vartheta)$ be an almost surely full-rank $k \times (k - p_\psi)$ matrix that spans the null-space of $\tilde{V}_f(\vartheta)^{-1/2} \hat{J}_T(\vartheta)$ such that

$$\hat{C}(\vartheta)\hat{C}(\vartheta)' = M_{\tilde{V}_f(\vartheta)^{-1/2}\hat{J}_t(\vartheta)} \tag{2.6}$$

$$M_x = (I - P_X), \quad P_X = X\left(X'X\right)X'. \tag{2.7}$$

2.1.2. Weak Identification Aspects

In particular, Martínez-Iriarte et al. (2020) develops an asymptotic theory framework based on fixed-smoothing asymptotics for the test statistics in order to account for the estimation uncertainty in the underlying LRV estimators. Consider the following long-run variance estimator

$$V_{ff}(\theta) = \lim_{T\to\infty} \mathsf{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(Y_t,\theta) \right). \tag{2.8}$$

Therefore, a non-parametric estimator of the LRV takes the quadratic form below

$$\hat{V}_{ff}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \omega_h\left(\frac{t}{T},\frac{t}{T}\right)\left[f(Y_t,\theta)-\bar{f}(Y_t,\theta)\right]\left[f(Y_t,\theta)-\bar{f}(Y_t,\theta)\right]' \tag{2.9}$$

$$\bar{f}(Y_t,\theta) = \frac{1}{T}\sum_{s=1}^{T} f(Y_t,\theta), \tag{2.10}$$

such that $\omega(.,.)$ is a weighting function, and h is the smoothing parameter indicating the amount of nonparametric smoothing. For example, we can estimate the kernel density the following way

$$\omega_h\left(\frac{t}{T},\frac{t}{T}\right) = k\left(\frac{(t-s)}{hT}\right) \tag{2.11}$$

for some kernel function $k(.)$, leading to the usual kernel LRV estimator. Thus, by substituting the smoothing estimator of the particular kernel function, we obtain the following test statistic

$$Q_T(\theta) = \frac{1}{2}\left[\frac{1}{\sqrt{T}}\sum_{s=1}^{T} f(Y_t,\theta)\right]' \hat{V}_{ff}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{s=1}^{T} f(Y_t,\theta)\right]. \tag{2.12}$$

Then, the K statistic is based on the first-order derivative of $Q_T(\theta)$. Define as below the gradients

$$\mathrm{g}_j(Y_t,\theta) = \frac{\partial f(Y_t,\theta)}{\partial \theta_j} \in \mathbb{R}^{m\times 1}, \;\; j\in\{1,...,d\}, \tag{2.13}$$

$$\mathrm{g}(Y_t,\theta) = \frac{\partial f(Y_t,\theta)}{\partial \theta'} = \left(\mathrm{g}_1(Y_t,\theta),....,\mathrm{g}_d(Y_t,\theta)\right) \in \mathbb{R}^{m\times d}, \tag{2.14}$$

$$\bar{\mathrm{g}}(Y_t,\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{t=1}^{T} \frac{\partial f(Y_t,\theta)}{\partial \theta'} \in \mathbb{R}^{m\times d}. \tag{2.15}$$

Taking the first-order and second-order derivatives of $\hat{V}_{ff}(\theta)$ with respect to θ_j, we obtain

$$\hat{V}_{\mathrm{g}_j f}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \omega_h\left(\frac{t}{T},\frac{t}{T}\right)\left[\mathrm{g}_j(Y_t,\theta)-\bar{\mathrm{g}}_j(Y_t,\theta)\right]\left[f(Y_t,\theta)-\bar{f}(Y_t,\theta)\right]' \tag{2.16}$$

$$\hat{V}_{\mathrm{g}_j \mathrm{g}_j}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \omega_h\left(\frac{t}{T},\frac{t}{T}\right)\left[\mathrm{g}_j(Y_t,\theta)-\bar{\mathrm{g}}_j(Y_t,\theta)\right]\left[\mathrm{g}_j(Y_t,\theta)-\bar{\mathrm{g}}_j(Y_t,\theta)\right]' \tag{2.17}$$

Then, it follows that

$$\frac{\partial Q_T(\theta)}{\partial \theta} = D_T(\theta) V_{ff}^{-1}(\theta) \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(Y_t, \theta) \right], \tag{2.18}$$

Denote with $D_T(\theta) = \left[D_{T,1}(\theta), ..., D_{T,d}(\theta) \right] \in \mathbb{R}^{m \times d}$, such that

$$D_{T,j}(\theta) = \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathrm{g}_j(Y_t, \theta) \right] - \hat{V}_{\mathrm{g},f}(\theta) \hat{V}_{ff}^{-1}(\theta) \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_j(Y_t, \theta) \right] \in \mathbb{R}^{m \times 1}. \tag{2.19}$$

Then, the K statistic for testing the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis given by $H_1 : \theta \neq \theta_0$ is given by

$$\mathscr{K}_T(\theta_0) = \left(\frac{\partial Q_T(\theta_0)}{\partial \theta} \right)' \left[D_T(\theta_0)' \bar{V}_{ff}^{-1}(\theta_0) D_T(\theta_0) \right]^{-1} \left(\frac{\partial Q_T(\theta_0)}{\partial \theta} \right). \tag{2.20}$$

where for any concave function $\phi(\theta)$, $\partial \phi(\theta_0) / \partial \theta$ is defined to be

$$\frac{\partial \phi(\theta_0)}{\partial \theta} = \left. \frac{\partial \phi(\theta)}{\partial \theta} \right|_{\theta = \theta_0} \in \mathbb{R}^{d \times 1}. \tag{2.21}$$

Thus, to consider fixed-smoothing asymptotics, we employ the orthonormal series LRV estimator

$$\omega_h \left(\frac{t}{T}, \frac{s}{T} \right) = \frac{1}{G} \sum_{\ell=1}^{G} \Phi_\ell \left(\frac{t}{T} \right) \Phi_\ell \left(\frac{s}{T} \right), \tag{2.22}$$

where G is a smoothing parameter for this estimator and $\Phi_\ell(.)$ is a set of a basis functions on $L^2[0,1]$. The weighting function is expressed with respect to a set of basis functions on the space of $L^2[0,1]$. Therefore, the LRV estimator takes the following form

$$\bar{V}_{ff}(\theta) = \frac{1}{G} \sum_{\ell=1}^{G} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_\ell \left(\frac{t}{T} \right) \left[f(Y_t, \theta) - \bar{f}(Y_t, \theta) \right] \right\} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_\ell \left(\frac{t}{T} \right) \left[f(Y_t, \theta) - \bar{f}(Y_t, \theta) \right] \right\}'.$$

Then, an updated estimator for $\mathscr{J}_T(\theta_0)$ needs to be obtained from the sample such that

$$\mathscr{J}_T(\theta_0) = \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{f}(Y_t, \theta_0) \right]' \hat{V}_{ff}^{-1}(\theta_0) \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{f}(Y_t, \theta_0) \right] \tag{2.23}$$

Remark 1. The modified statistic is not the same as the original statistic due to the projection of the function into the space which is induced by the transformation of the column vector space. This property allows us to obtain a consistent estimator $\hat{\theta}$ of θ_0. However, to obtain an unbiased estimator for the variance of the estimator, we also need to obtain unbiased estimators for each partial derivative that the covariance matrix is composed to.

Specifically, for the variance estimator we obtain the following

$$V(\theta_0) := V = \begin{bmatrix} V_{ff}(\theta_0) & V_{fg}(\theta_0) \\ V_{gf}(\theta_0) & V_{gg}(\theta_0) \end{bmatrix}. \tag{2.24}$$

Therefore, the CLT to hold the following asymptotic distribution to hold

$$\begin{pmatrix} \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{T} \left\{ f(Y_t, \theta_0) - \mathbb{E}\big[f(Y_t, \theta_0) \big] \right\} \\ \\ \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{T} \mathsf{vec}\left\{ \mathrm{g}(Y_t, \theta_0) - \mathbb{E}\big[\mathrm{g}(Y_t, \theta_0) \big] \right\} \end{pmatrix} \Rightarrow \begin{pmatrix} \psi_f \\ \psi_{\mathrm{g}} \end{pmatrix}, \tag{2.25}$$

where $\psi_f \in \mathbb{R}^{m \times 1}$ and $\psi_{\mathrm{g}} \in \mathbb{R}^{md \times 1}$. Therefore, it holds that we have a sequence of matrices

$$T^{\kappa} \times \mathbb{E}\left[\mathrm{g}(Y_t, \theta_0) \right] \to \Pi = (\Pi_1, ..., \Pi_d) \in \mathbb{R}^{m \times d} \tag{2.26}$$

In other words, we consider the convergence rate of the $m-$system equations. By multiplying with T^{κ} we ensure that we take into account the different convergence rate depending on the type of functional form specification for the model under investigation. Specifically, when $\kappa = 0$, then the $m-$moment conditions doesn't include the correct rate of convergence which implies that the matrix Π has a full column rank and therefore the parameter under the null hypothesis θ_0, can be estimated at the usual parametric $\sqrt{T}$-rate. In other words, the case which corresponds to the weak identification of the model specification occurs when $\kappa = 1/2$, since it asymptotically converge into the null matrix, such that, $\Pi = 0$ and therefore, θ_0 cannot be consistently estimated (see, Martínez-Iriarte et al. (2020)). Therefore, the estimation procedure for the case of fixed autocorrelation is given as below

$$D_{T,j}(\theta_0) - \sqrt{T}\mathbb{E}\left[\mathrm{g}(Y_t, \theta_0) \right] = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ \mathrm{g}(Y_t, \theta_0) - \mathbb{E}\big[\mathrm{g}(Y_t, \theta_0) \big] \right\} - \hat{V}^{-1}_{\mathrm{g}_j, f}(\theta_0) \hat{V}^{-1}_{ff}(\theta_0) \times \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(Y_t, \theta_0) \right]$$

Now, to estimate the above moment conditions the important component of the estimation procedure is to obtain unbiased estimators for the LRV covariance matrices which are computed based on a set of basis functions. Let $\ell \in \{1, ..., G\}$, then the basis functions shall satisfy: *(i)*. $\Phi_\ell(.)$ are piesewise monotonic, continuously differentiable, and *(ii)*. $\Phi_\ell(.)$ are orthonormal in the space of $L^2[0,1]$ functions and satisfy $\int_0^1 \Phi_\ell(x) dx = 0$. Therefore, the corresponding estimators are obtained as below

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_\ell \left(\frac{t}{T} \right) \left[f(Y_t, \theta_0) - \bar{f}(Y, \theta_0) \right] \Rightarrow \int_0^1 \Phi_\ell(r) dB_f(r) \tag{2.27}$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_\ell \left(\frac{t}{T} \right) \left[\mathrm{g}(Y_t, \theta_0) - \bar{\mathrm{g}}(Y, \theta_0) \right] \Rightarrow \int_0^1 \Phi_\ell(r) dB_f(r) \tag{2.28}$$