1 GMM: a brief introduction

In the context of the linear model of Instrumental Variables, we observed that the statistical model was specified through conditions involving equalities of expected values of functions of the sample. This concept is closely tied to semiparametric models; for instance, if we are only interested in the mean of D_0, and we are somehow averse to the risk of misspecification, it is not necessary to provide a parametric specification for the entire distribution.[1]

The Instrumental Variables Estimator (IVE) essentially arose from a relevant *analogy principle* given the population moment conditions; the moments with respect to D_0 were approximated by their empirical counterparts, and the system was solved with respect to θ, or more generally, an element that minimizes some norm of the vector of those empirical moments was found. This is a potentially more complex version of a relatively straightforward methodology: if we are interested in the mean of D_0 and have access to a sample of random variables following it, we can approximate the unknown mean using the sample mean among other methods. This methodology constitutes the Method of Moments (MoM), a semiparametric method that broadly estimates parameters related to moments using their empirical counterparts-see, for example, [3]. The dimension of the parameters involved equals the number of moments used.

The Generalized Method of Moments (GMM) generalizes this approach by allowing the number of moments to be greater than or equal to the

[1] This is related to the moment problem in probability theory; given a list of moments, how can we determine the set of distributions that satisfy them.

dimension of the parameter vector, thereby enriching the initial theory. This, among other things, allows for the use of more information to identify the latent θ_0. This generalization was introduced by L. P. Hansen in [7] and is particularly useful in Econometrics, where Economic Theory and Empirical Economics often prescribe "high"-dimensional semiparametric specifications.

This methodology also encompasses statistical inference procedures derived from optimization, where first-order conditions take the form of expected value equalities. For instance, approximations of Maximum Likelihood Estimators (MLE) or of Quasi-Likelihood Estimators-can be considered estimators adherent the method (GMME)-see the concept of the score estimator.

Since the dimension of the moments vector does not necessarily match the dimension of the parameter, extracting the GMME involves minimizing the norm of the former with respect to the latter. We have already seen this in the case of IVE. GMME is a special case of the broader framework of optimization-based statistical inference procedures previously discussed. In the following, we describe the relevant background[2] and integrate it into the previous framework regarding the definition and properties of GMME. Consequently, constructing and describing the tests mentioned in the previous section becomes straightforward and is not repeated. Instead, we develop a specification test for the model using reasoning similar to that in the previous section. In any case, a prototypical example of the following is the already developed model of Instrumental Variables.

Using the general notation developed so far, let $m : \mathbb{R}^k \times \Theta \to \mathbb{R}^q$ be

[2]In a relatively general form but not the most general possible.

appropriately measurable, with $q \geq p$, and the model specification arises from the moment vector $\mathbb{E}(m(z, \theta_0)) = \mathbf{0}_q$, where integration is with respect to D_0, assuming that the distribution of the ith k-dimensional sample element is independent of i.[3] This forms-based on the above-the model $\{\mathbb{E}(m(z, \theta)) = \mathbf{0}_q,\ \theta \in \Theta\}$, which is interpreted according to the previously established statistical model interpretations. Thus, within the Instrumental Variables framework, in our current context, we have $m(z_i, \theta) = \mathbf{W}_i'(Y_i - \mathbf{X}_i\theta)$, where z_i is the ith k-dimensional sample element, $\mathbf{W}_i$ is the ith row of $\mathbf{W}_n$, $\mathbf{X}_i$ is the ith row of $\mathbf{X}_n$, and Y_i is the ith element of Y_n.

To complete the background and construct the relevant objective function c_n, we will take a slightly more general approach compared to the methodology seen in the example of instrumental variables-although the path we follow is outlined in related exercises. In this context, the matrix determining the norm will be allowed to be stochastic and/or depend on n. This will later enable reasoning about the statistical approximation of its optimal choice. Thus, $\mathbf{W}_n$ is, for almost every possible sample value, a strictly positive definite $q \times q$ matrix. In this theory, $\mathbf{W}_n$ is typically referred to as a weighting matrix. Given this, we consider the objective function:

$$c_n(\theta, \mathbf{W}_n) := \frac{1}{n^2} \left(\sum_{i=1}^{n} m(z_i, \theta) \right)' \mathbf{W}_n \left(\sum_{i=1}^{n} m(z_i, \theta) \right), \tag{1}$$

which is essentially the square of the stochastic norm $\mathbb{R}^q \ni \mathbf{x} \to \mathbf{x}'\mathbf{W}_n\mathbf{x}$.

[3]This holds, for example, in iid settings or stationary time series environments. It nevertheless can be generalized, yet at the cost of a more complex presentation. For instance, the distribution over which the integration is performed could be appropriately conditional, and m could depend on i, etc. We will not deal with such generalizations for simplicity.

Squaring is used as it enables properties such as differentiability and strict convexity-in at least some cases, which, as we know from our general theory, can facilitate the study of the estimator's properties. The square is evaluated based on the sample analogue of the expected value of the moment vector, which in turn is computed at θ. The notation involving the weighting matrix emphasizes the dependence of the examined objects on it. We will see below that this is reasonable.

Given the objective function, and some related optimization error, the GMME is the relevant extremum estimator (OE):

$$c_n(\theta_n(\mathbf{W}_n), \mathbf{W}_n) \leq \inf_{\theta \in \Theta} c_n(\theta, \mathbf{W}_n) + u_n. \tag{2}$$

Conditions for existence are described in the relevant part of our general theory. For example, when m is continuous with respect to θ for almost every possible sample value and the parameter space is compact, the estimator exists. In general, the estimator will also depend on the weighting matrix.

> *Exersice*: Can you describe sufficient conditions under which the estimator is independent of weighting?

Our work in the previous parts of the notes directly indicates that both the IVE and the OLSE are examples of GMME, with the latter notably not depending on the weighting matrix.

For reasons already explained in constructing the general theory of extremum estimators, we proceed to the asymptotic properties of the GMME. These will arise from the specialization of our general asymptotic theory.

1.1 Weak Consistency

According to what is outlined in the general theory of OE, consistency will arise through the appropriate convergence of c_n to a suitable objective function that is uniquely minimized at θ_0.

The form of the objective function suggests that the limit we seek will have the form of a quadratic expression evaluated at $\mathbb{E}(m(z,\theta))$, as this is appropriately approximated by its empirical analogue $\frac{1}{n}\sum_{i=1}^{n} m(z_i,\theta)$. The quadratic form will be shaped by the asymptotic behavior of $\mathbf{W}_n$. If this is appropriately designed so that the limiting function is strictly convex, then asymptotic identification will be ensured provided the moment vector $\mathbb{E}(m(z,\theta))$ vanishes only at θ_0. One way to obtain such a limiting objective function is through the following high-order conditions.

Assumption 1. *The following hold:*

(A). $m(z,\theta)$ *is continuous with respect to* θ*, almost for every sample value,*

(B). $\frac{1}{n}\sum_{i=1}^{n} m(z_i,\theta) \overset{cp}{\to} \mathbb{E}(m(z,\theta))$,

(C). There exists a strictly positive definite non-stochastic matrix W *such that* $W_n \overset{p}{\to} W$*, and*

(D). $\mathbb{E}(m(z,\theta)) = \mathbf{0}_q \Leftrightarrow \theta = \theta_0$.

Condition (A) ensures the continuity of c_n almost surely for every sample value. Condition (B) can arise, for instance, in iid settings or under stationarity and ergodicity, through suitable uniform LLNs or pointwise LLNs combined with some strong property of "joint" (with respect to z) continuity for $\|m(z,\theta)\|$ with respect to θ.Condition (C) can similarly arise, and in combination with the consistency of some initial estimator for θ_0 on which it may depend-we will examine this in more detail below. The limiting matrix could also be stochastic for the following result. Conditions (B) and (C) imply that $c_n \overset{\text{cp}}{\to} c := (\mathbb{E}(m(z,\theta)))' W(\mathbb{E}(m(z,\theta)))$. This is because, for arbitrary θ and $\theta_n^\star \to \theta$, due to the triangle inequality,

$$|c_n(\theta_n, W_n) - c(\theta, W)| \leq$$

$$|c_n(\theta_n, W_n) - c_n(\theta, W)| + |c_n(\theta_n, W) - c(\theta, W)|.$$

The first term in the previous inequality, due to a submultiplicative property of a suitable matrix norm and related properties of the Cholesky factorization,[4] is less than or equal to $\|\frac{1}{n}\sum_{i=1}^{n} m(z_i,\theta) - \mathbb{E}(m(z,\theta))\|^2 \|W_n - W\|$. Due to (B), (C), and the Continuous Mapping Theorem , this converges in probability to zero. Similarly, the second term above is less than or equal to $\|\frac{1}{n}\sum_{i=1}^{n} m(z_i,\theta) - \mathbb{E}(m(z,\theta))\|^2 \|W\|$, which also converges in probability to zero due to (B). Conditions (C)-(D) complete the background required for the application of Consistency Theorem in our general theory, ensuring that c is uniquely minimized at θ_0 (why?):

Theorem 1. *Let Θ be compact, $u_n = o_p(1)$, and assume that Assumption 1 holds. Then the GMME is a weakly consistent estimator of θ_0.*

> *Exersice*: Under what conditions would the application of the theorem

[4] We refer to properties of the Frobenius norm, which essentially extends the Euclidean norm to matrix spaces-the interested reader may refer to ??? for further details.

that derives consistency from pointwise convergence to a strictly convex criterion be ensured?

For the application of the above to the case of IVE, see the related examples and exercises in previous sections of the notes.

1.2 Rate of Convergence and Asymptotic Distribution

Following the previous section, we will aim to describe sufficient conditions for applying our general theory regarding the determination of the rate of convergence and the asymptotic distribution of the GMME.

We will focus on the first two sufficient conditions for our general results to hold, as the others are not essentially related to the form of the objective function, given its domain. We will assume that m is twice continuously differentiable in a neighborhood of θ_0, and the derivatives will be appropriately one-sided if θ_0 is a boundary point, possessing properties such that the local Taylor expansion of m to the second degree is valid.[5] Since the matrix W_n is constructed to be independent of θ, this means, in the notation of our general theory,

$$g(\theta_0) = 2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m'(z_i,\theta_0)}{\partial\theta}\right)W_n\left(\frac{1}{n}\sum_{i=1}^{n}m(z_i,\theta_0)\right),$$

and

$$\begin{aligned} q(\theta_n^{\star\star}) &= 2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m'(z_i,\theta_n^{\star\star})}{\partial\theta}\right)W_n\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\theta_n^{\star\star})}{\partial\theta'}\right) \\ &+2\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m'(z_i,\theta_n^{\star\star})}{\partial\theta\partial\theta_j}\right)W_n\frac{1}{n}\sum_{i=1}^{n}m(z_i,\theta_n^{\star\star})\right]_{j=1,\ldots,p}. \end{aligned}$$

[5]The interested reader is referred to [1] for a general formulation of such an assumption.

These expressions guide us in establishing the validity of the second part of the assumption of our general theory. Without addressing the most general case, we limit ourselves to the following assumption:

Assumption 2. *i. The term $\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\theta)}{\partial\theta'}$ converges continuously in probability to a non-stochastic matrix, say ∂m_{θ_0}, of dimensions $q\times p$, with rank p, at θ_0.*

ii. For any $\theta_n^{\star}\to\theta_0$, $\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m'(z_i,\theta_n^{\star})}{\partial\theta\partial\theta_j}\| = O_p(1)$, for all $j=1,\ldots,p$.

iii. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}m(z_i,\theta_0)\rightsquigarrow z\sim N(\mathbf{0}_q,V_{m,\theta_0})$, where the $q\times q$ matrix V_{m,θ_0} is positive definite.

The convergences in 2.(i),(ii) can be satisfied via locally uniform LLNs in iid or stationary and ergodic settings. The convergence in 2.(iii) can be satisfied via a Central Limit Theorem like the one we have examined previously, provided the appropriate conditions of dependence and existence of moments hold. Assumption 2.(ii) ensures that, for any related sequence, the term $\|\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 m'(z_i,\theta_n^{\star\star})}{\partial\theta\partial\theta_j}\right)W_n\frac{1}{n}\sum_{i=1}^{n}m(z_i,\theta_n^{\star\star})\right]_{j=1,\ldots,p}\|$ converges in probability to 0, due to 1. Similarly, 2.(i) ensures, together with 1.C and the Continuous Mapping Theorem , the convergence in probability of the quadratic form $(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m'(z_i,\theta_n^{\star\star})}{\partial\theta})W_n(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\theta_n^{\star\star})}{\partial\theta'})$ to $\partial m'_{\theta_0}W\partial m_{\theta_0}$, and this $p\times p$ matrix has rank p. Therefore, the second part of the general assumption in our general theory holds with $\breve{\mathbf{J}}_{\theta_0}:=\partial m'_{\theta_0}W\partial m_{\theta_0}$. Consequently, 2.(i), (iii), 1.C, Sluysky's Lemma, and the Continuous Mapping Theorem ensure that for $r_n=\sqrt{n}$, we have the Gaussian random vector $z_{\theta_0}=2\partial m'_{\theta_0}Wz\sim N(\mathbf{0}_p,4\partial m'_{\theta_0}WV_{m,\theta_0}W\partial m_{\theta_0})$. In conclusion:

Theorem 2. *If Assumptions 1, 2, and the third and fourth parts of our general assumption hold, then the conclusions of our general theorem apply to the GMME with* $r_n = \sqrt{n}$, $\breve{\mathbf{J}}_{\theta_0} := \partial m'_{\theta_0} W \partial m_{\theta_0}$, *and* $z_{\theta_0} = 2\partial m'_{\theta_0} W z \sim N(\mathbf{0}_p, 4\partial m'_{\theta_0} W V_{m,\theta_0} W \partial m_{\theta_0})$.

When θ_0 lies in the interior of Θ, the above can be interpreted as

$$\sqrt{n}(\theta_n(W_n) - \theta_0) \rightsquigarrow N(\mathbf{0}_p, V_{\theta_0}),$$

where $V_{\theta_0} := (\partial m'_{\theta_0} W \partial m_{\theta_0})^{-1} \partial m'_{\theta_0} W V_{m,\theta_0} W \partial m_{\theta_0} (\partial m'_{\theta_0} W \partial m_{\theta_0})^{-1}$.

In any case, regardless of the location of the true parameter value, we observe that when $p = q$, the random element $-\frac{1}{2}\breve{\mathbf{J}}_{\theta_0} z_{\theta_0}$, which determines the asymptotic form of the estimator, follows $N(\mathbf{0}, ((\partial m_{\theta_0})^{-1})' V_{m,\theta_0} (\partial m_{\theta_0})^{-1})$. This occurs because all involved matrices are square and invertible, and due to the properties of the inverse of a product and the commutativity between inversion and transposition. However, this is independent of the weighting matrix, so:

Lemma 1. *When* $p = q$*, the GMME is asymptotically independent of weighting.*

In the case where $q > p$, asymptotic independence does not generally hold. This raises the question of the optimal choice of weighting, such that the resulting W yields the optimal asymptotic variance, regardless of the location of the true parameter value. We will address this in the next section.

Exersice: Specify the above in the case of the IVE. Relevant exercises from the previous sections may be taken into account.

1.3 Optimal Choice of Weighting

From the above, the issue of optimal selection of W_n arises to minimize asymptotic variance.[6][7] Based on the above results, this is trivial in the case where $p = q$: in this case, any choice is optimal. But what happens when there are more integrals in the moment condition than components in the parameters?

We observe that for $W = V_{m,\theta_0}^{-1}$, the asymptotic variance of $-\frac{1}{2}\mathbf{J}_0 z_{\theta_0}$ becomes $(\partial m'_{\theta_0} V_{m,\theta_0}^{-1} \partial m_{\theta_0})^{-1}$.[8] It is proven that this corresponds to the asymptotically optimal choice. This is because, since V_{m,θ_0} is positive definite, it can be factored as a product LL', where L is an appropriate triangular matrix. Using this factorization, it follows that (**Exercise**: perform the calculations in detail):

$$(\partial m'_{\theta_0} W \partial m_{\theta_0})^{-1} \partial m'_{\theta_0} W V_{m,\theta_0} W \partial m_{\theta_0} (\partial m'_{\theta_0} W \partial m_{\theta_0})^{-1} - (\partial m'_{\theta_0} V_{m,\theta_0}^{-1} \partial m_{\theta_0})^{-1} = KLL'K'$$

where $K := (\partial m'_{\theta_0} W \partial m_{\theta_0})^{-1} \partial m'_{\theta_0} W - (\partial m'_{\theta_0} V_{m,\theta_0}^{-1} \partial m_{\theta_0})^{-1} \partial m'_{\theta_0} V_{m,\theta_0}^{-1}$. The matrix

[6]This pertains to ordering based on the positive definiteness of the difference. As previously mentioned, the optimal choice will, by definition, have the property that the difference between any other asymptotic variance and the optimal one will be positive definite.

[7]It is noted that in what follows, the moment vector m is assumed given. We do not address the equally important issue of the optimal selection of a moment vector when there is availability-a problem that also relates to geometric characteristics of the distributions forming the respective statistical model.

[8]We note that this collapses to the expression obtained in the case where dimensions match.

$KLL'K' = (KL)(KL)'$ is necessarily semi-positive definite.

The above, combined with Assumption 1.C, implies that any choice of W_n such that it converges in probability to V_{m,θ_0}^{-1} yields the asymptotically efficient estimator given the moment vector. Given Assumption 2.(iii), any consistent estimator of the asymptotic variance of the empirically weighted moments' vector scaled by $\sqrt{n}$ and evaluated at θ_0, i.e., $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i, \theta_0)$, leads, through its inversion, to a consistent estimator of the optimal weighting due to the Continuous Mapping Theorem . Generally, this asymptotic variance-and therefore the optimal weighting-will depend on the unknown θ_0 (**Exercise**: Is this true for the case of the GMME?). In this case, constructing an asymptotically efficient estimator for the variance can be facilitated if (a) there exists a random element $V_n(\theta)$, a function of the sample and the parameter, that converges continuously in probability to V_{m,θ_0} at θ_0, and (b) an initial consistent estimator of θ_0, say $\theta_n^{\triangle}$, is available. Under (a) and (b), through the invoked mode of convergence, $V_n(\theta_n^{\triangle})$ will be a weakly consistent estimator of the asymptotic variance, and consequently, $V_n^{-1}(\theta_n^{\triangle})$ will be a weakly consistent estimator of the asymptotically optimal weighting.

The estimator $\theta_n^{\triangle}$ can be chosen as $\theta_n(W)$, where W is an arbitrary weighting matrix, independent of θ_0, e.g., $W = \mathbf{I}_q$.[9] In this case, the GMME $\theta_n(V_n^{-1}(\theta_n(W)))$ is obtained through a two-step optimization process: in the first step, $\theta_n(W)$ is calculated, and in the second, $\theta_n(V_n^{-1}(\theta_n(W)))$ is derived. This estimator is called the 2-GMME. Clearly, starting from W, this process can be extended to an arbitrary number of steps: if $m \geq 2$, the m-GMME $\theta_n^{(m)}$ is defined as $\theta_n(V_n^{-1}(\theta_n^{(m-1)}))$, with $\theta_n^{(1)} := \theta_n(W)$.

Another variant, which does not initially require an estimation of θ_0, al-

[9]Based on our assumptions, consistency does not require an optimal choice of weighting.

lows V_n to depend freely on θ within the criterion being optimized. This leads to the continuously updated estimator (CUE), defined by the approximate minimization:

$$\frac{1}{n^2}(\sum_{i=1}^{n} m(z_i, \theta_n^{\text{CUE}}))' V_n^{-1}(\theta_n^{\text{CUE}})(\sum_{i=1}^{n} m(z_i, \theta_n^{\text{CUE}}))$$

$$\leq \inf_{\theta\in\Theta} \frac{1}{n^2}(\sum_{i=1}^{n} m(z_i, \theta))' V_n^{-1}(\theta)(\sum_{i=1}^{n} m(z_i, \theta)) + u_n.$$

Try to extend Assumptions 1 and 2 appropriately to show that the 2-GMME, m-GMME, and CUE are asymptotically equivalent-i.e., they fall under Theorem 2 with the same, optimal asymptotic variance.[10]

The form that $V_n(\theta)$ may take is partly determined by Assumption 2.(iii) and strongly depends on the properties of the sequence $(m(z_i, \theta))_{i\in\mathbb{N}}$. For instance, in an iid setting, or more generally in a stationary and ergodic setting where additionally the conditional covariances appearing in the CLT we have examined in previous section of the notes are zero, a property of uniform integrability that holds continuously at θ_0 can show that the stochastic matrix $\frac{1}{n}\sum_{i=1}^{n} m(z_i, \theta)m'(z_i, \theta)$ converges continuously in probability to V_{m,θ_0} at θ_0, and thus $(\frac{1}{n}\sum_{i=1}^{n} m(z_i, \theta_n^{\triangle})m'(z_i, \theta_n^{\triangle}))^{-1}$ is a consistent estimator of the optimal weighting.

When this is not the case, the above must be replaced by an estimator that approximates the mode at which the asymptotic variance is affected by the dependence between $m(z_i, \theta)$ and $m(z_j, \theta)$, for $i \neq j$ as $n \to \infty$.[11]

[10] However, they may differ in interesting ways with respect to finer asymptotic properties that are not examined in this book. The interested reader is referred, for example, to [8].

[11] The interested reader may consult [2] for a related methodology that applies in time-

We conclude this section with the issue of consistently estimating the optimal asymptotic variance. This may be of interest, among other things, for the application of hypothesis testing procedures such as those developed in the asymptotic testing theory we have examined and the exercises therein. Due to Assumption 2.(i) and the construction above, it follows via the Continuous Mapping Theorem (**Exercise**: explain the details!) that

$$((\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m'(z_i,\theta_n^{\triangle})}{\partial\theta})V_n^{-1}(\theta_n^{\triangle})(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\theta_n^{\triangle})}{\partial\theta'}))^{-1}$$

is a consistent estimator of the optimal asymptotic variance that arises in our framework.

1.4 A Test for Partial Specification

In this section, we will challenge the basic assumption of correct model specification that lies at the core of our investigations. The condition $\mathbb{E}(m(z,\theta_0)) = \mathbf{0}_q$ is inherent in the specification of the model. Is it possible to use the asymptotically optimal GMME to statistically test whether this condition is ultimately correct? Clearly, we are dealing with a hypothesis structure that does not fall under the ones which we addressed in our general framework. The current hypotheses can be described as follows:

$$\begin{aligned} H_0 &: \ \mathbb{E}(m(z,\theta)) = \mathbf{0}_q, \ \exists!\theta \in \Theta \\ H_1 &: \ \mathbb{E}(m(z,\theta)) \neq \mathbf{0}_q, \ \forall\theta \in \Theta. \end{aligned} \tag{3}$$

series environments and addresses a broad range of types of temporal dependence.

Under the null hypothesis, there exists a point in the parameter space that satisfies the system of equality constraints-under the identification condition used this will necessarily be unique. Under the alternative hypothesis, there will be no such point, either because one does not exist in general in $\mathbb{R}^p$, in which case the specification error is global,[12] or because θ_0 has been incorrectly excluded via an inappropriate choice of the parameter space Θ. Note that the asymptotic theory we have outlined does not generally hold under the alternative hypothesis. The limit theory could however be modified to describe the limiting behavior of the estimator under the alternative hypothesis. This modification is not very involved in the second case of specification error.

For the design of a testing procedure for the above hypotheses, we will focus on the asymptotic behavior of the random vector of empirical moments evaluated at the estimator[13] $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i, \theta_n)$ under Assumptions 1, and 2 and correct specification. We observe that due to the assumed differentiability and given that θ_0 is an interior point, we will have that, with probability converging to 1,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i, \theta_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i, \theta_0) + \frac{1}{n}\sum_{i=1}^{n} \frac{\partial m(z_i, \theta_n^{\star})}{\partial \theta'})\sqrt{n}(\theta_n - \theta_0),$$

where, as is known, $\theta_n^{\star}$ lies on the line segment connecting the estimator to θ_0, and thus, due to Theorem 1, will converge in probability to θ_0 (why?).

[12]That is, the specific system of moments is not satisfied.

[13]For simplicity, the dependence of the estimator on the weighting matrix will not be explicitly stated.

This, combined with Assumption 2.(i), implies that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\theta_n^{\star})}{\partial\theta'}) \xrightarrow{\text{p}} \partial m_{\theta_0}.$$

Additionally, Assumption 2.(iii), combined with Theorem 2, implies the joint convergence

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i,\theta_0), \sqrt{n}(\theta_n-\theta_0)\right) \rightsquigarrow \left(z, -\partial m_{\theta_0}(\partial m_{\theta_0}' V_{m,\theta_0}^{-1}\partial m_{\theta_0})^{-1}\partial m_{\theta_0}' V_{m,\theta_0}^{-1} z\right).$$

This, along with Slutsky's Lemma and the Continuous Mapping Theorem , ultimately implies that when the model is well-specified and under 1 and 2,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(z_i,\theta_n) \rightsquigarrow N(\mathbf{0}_q, (\mathbf{I}_q - M_{\theta_0})V_{m,\theta_0}(\mathbf{I}_q - M_{\theta_0})'), \tag{4}$$

where

$$M_{\theta_0} := \partial m_{\theta_0}(\partial m_{\theta_0}' V_{m,\theta_0}^{-1}\partial m_{\theta_0})^{-1}\partial m_{\theta_0}' V_{m,\theta_0}^{-1}.$$

> *Exersice*: Show that M_{θ_0} is idempotent.[a] Then show that $\mathbf{I}_q - M_{\theta_0}$ is idempotent.
>
> [a]That is, it satisfies $M_{\theta_0}^2 = M_{\theta_0}$.

> *Exersice*: Show that $\text{tr}(M_{\theta_0}) = p$. Then show that $\text{tr}(\mathbf{I}_q - M_{\theta_0}) = q - p$.

The above exercises outline the proof of:

Lemma 2. *rank*$((\mathbf{I}_q - M_{\theta_0})V_{m,\theta_0}(\mathbf{I}_q - M_{\theta_0})') = q - p$.

Proof. Since $\mathbf{I}_q - M_{\theta_0}$ is idempotent, its eigenvalues can only take the values 0 or 1. By construction, their sum equals the trace of the matrix, which is $q - p$. Therefore, the matrix necessarily has $q - p$ unit eigenvalues and p zero eigenvalues. Consequently, its rank equals $q - p$, and this is inherited (why?) by the rank of $(\mathbf{I}_q - M_{\theta_0})V_{m,\theta_0}(\mathbf{I}_q - M_{\theta_0})'$. □

When $p = q$, $(\mathbf{I}_q - M_{\theta_0})V_{m,\theta_0}(\mathbf{I}_q - M_{\theta_0})'$ is zero, so 4 describes a non-degenerate matrix only when $p > q$. We therefore proceed by focusing exclusively on the case where $p > q$. The above are not useful in cases where the number of parameter components matches the number of conditions.

The construction of the desired test may be facilitated if-using the above asymptotic theory-we manage to construct a test statistic with a quadratic form, derive a procedure that has an asymptotic chi-squared distribution under the null hypothesis, and proceed as in our general OE theory to determine the asymptotic rejection region. The problem is that even in the case where $p < q$, the above asymptotic distribution is partially degenerate since it assigns zero probability to a p-dimensional subspace of $\mathbb{R}^p$; its variance is not an invertible matrix.

To construct the test, as outlined earlier, we need the following concept of the generalized inverse matrix, based on which the required quadratic form will be constructed:

> If A is an $m \times n$ matrix, then the Moore-Penrose generalized inverse of A, denoted A^+, is defined as any $n \times m$ matrix satisfying:[a]
>
> $$(i).\ AA^+A = A,\ (ii).\ A^+AA^+ = A^+,\ (iii).\ (AA^+)' = AA^+,$$

$$(iv). (A^+A)' = A^+A, \ (v). \text{rank}(A) = \text{rank}(A^+).$$

It can be shown-see Theorem 2.1 in [5]-that a matrix satisfying the above always exists and is unique. For example, if $A = \mathbf{0}_{m\times n}$, then $A^+ = \mathbf{0}_{n\times m}$, or if $A = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$ with $x_1 \neq 0$, then $A^+ = \left(\frac{1}{x_1}, 0\right)$. It is clear that if A is invertible, then its unique generalized inverse is the usual inverse. It is also shown-see Theorem 3.2 in the aforementioned work-that if $\mathbf{x} \sim N(\mathbf{0}_q, V)$ and $0 < \text{rank}(V) \leq q$,[b] then $\mathbf{x}'V^+\mathbf{x} \sim \chi^2_{\text{rank}(V)}$, which generalizes a previous similar observation in our examination of the behavior of quadratic forms w.r.t. Gaussian random vectors.

[a] Any matrix satisfying (i) is called a generalized inverse. It can be shown that the generalized inverse is not generally unique. This is because such a matrix is restricted to act as an isomorphism between $\text{im}(A)$ and $\mathbb{R}^m/\text{ker}(A)$, while acting arbitrarily on $\mathbb{R}^n - \text{im}(A)$. The remaining conditions enforce uniqueness, which arises via solving an optimization problem. For $\mathbf{y} \in \mathbb{R}^n - \text{im}(A)$, $A^+\mathbf{y}$ is obtained by finding the unique projection of $\mathbf{y}$ onto the closed and convex $\text{im}(A)$ and then inverting the projection through the invertible part of A.

[b] The comment on non-uniqueness tells us that $\text{rank}(A^+A) = \text{rank}(A)$. Also, from (i), we obtain $A^+AA^+A = A^+A$, so A^+A is idempotent.

From the above, the construction of a consistent estimator for the optimal weighting matrix described in Section 1.3, Slutsky's Lemma, and the Continuous Mapping Theorem , it follows-using the objects from the aforementioned section-that:

Lemma 3. *Under Assumptions 1 and 2, if the null hypothesis in 7 holds, and if θ_0 is in the interior of Θ, then for the random variable*

$$\mathcal{J}_n := \frac{1}{n}(\sum_{i=1}^{n} m(z_i, \theta_n))'((\mathbf{I}_q - M_n)V_n(\mathbf{I}_q - M_n)')^+(\sum_{i=1}^{n} m(z_i, \theta_n)),$$

with[14]

$$M_n := (\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\theta_n)}{\partial \theta'})\times$$
$$((\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m'(z_i,\theta_n)}{\partial \theta})V_n^{-1}(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i,\theta_n)}{\partial \theta'}))^{-1}\times$$
$$(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m'(z_i,\theta_n)}{\partial \theta})V_n^{-1},$$

and V_n *the respective weakly consistent estimator*[15] *of* V_{m,θ_0}*, it holds that:*

$$\mathcal{J}_n \rightsquigarrow \chi^2_{q-p}. \tag{5}$$

Proof. The above partially demonstrate the validity of 5. The uniqueness of the Moore-Penrose inverse, combined with the fact that the rank of $(\mathbf{I}_q - M_n)V_n(\mathbf{I}_q - M_n)'$ necessarily matches that of $(\mathbf{I}_q - M_{\theta_0})V_{m,\theta_0}(\mathbf{I}_q - M_{\theta_0})'$ (Exercise: prove this!), can be shown to imply that the first converges in probability to the second.[16] Thus, 5 ultimately follows. □

The above allows us to construct the following hypothesis test, commonly referred to as the J-test:

[14]Note that given the extraction of the estimator, there is no need to use the auxiliary $\theta_n^{\triangle}$-why?
[15]See Section 1.3.
[16]The interested reader may consult Corollary 8 in [12], or think in terms of the convergence of solutions in sequences of strictly convex optimization problems, similarly to our relevant general consistency theorem regarding convex criteria!

Algorithm 1 J-Test

1. Choose a significance level $\alpha \in (0, 1)$.
2. Given α, find

$$q_\alpha := \inf \left\{ x \in (0, +\infty) : \int_0^x \frac{z^{\frac{q-p}{2}-1}}{2^{q-p/2}\Gamma(\frac{q-p}{2})} \exp{(-\frac{z}{2})} dz \geq 1 - \alpha \right\}.$$

3. Define the rejection region for H_0 as the interval $(q_{1-\alpha}, +\infty)$.
4. Evaluate $\mathcal{J}_n$ on the sample value and reject H_0 if and only if $\mathcal{J}_n \in (q_{1-\alpha}, +\infty)$.

Based on the Lemma 3, it is easily proven:

Theorem 3. *Under Assumptions 1 and 2, if the null hypothesis in 7 holds, and if θ_0 is in the interior of Θ, then the J-test is asymptotically exact.*

> *Exersice*: How does the above theorem change if θ_0 lies on the boundary of the parameter space?

> *Exersice*: What is the form of the hypotheses structure anf of $\mathcal{J}_n$ in the case of IVE?

When the null hypothesis does not hold, we cannot say much about the asymptotic behavior of $\mathcal{J}_n$ unless further assumptions are made about the specification error of the model and the asymptotic behavior of the estimator under this error. For example, if the conditions are not globally misspecified but there exists a **unique** $\theta_0 \in \mathbb{R}^p - \Theta$ that satisfies them, then under Assumption 1.A, B, and provided that Θ is closed and convex, and the limiting criterion is strictly convex, it can be shown that $\theta_n \xrightarrow{\mathrm{p}} \theta_{00}$, a unique element of Θ. If in Assumption 2.(i), (ii), θ_0 is replaced with θ_{00}, Assumption 2.(iii) is modified to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m(z_i,\theta_{00}) - \mathbb{E}(m(z_0,\theta_{00}))) \rightsquigarrow z \sim N(\mathbf{0}_q, V_{m,\theta_{00}}),$$

with the $q \times q$ matrix $V_{m,\theta_{00}}$ strictly positive definite, and $\mathbb{E}(m(z_0,\theta_{00}))$ not in the kernel of the corresponding matrix $(\mathbf{I}_q - M_{\theta_{00}})V_{m,\theta_{00}}(\mathbf{I}_q - M_{\theta_{00}})'$, then $\mathcal{J}_n$ diverges to $+\infty$ (**Exercise**: prove this!), and thus the procedures is asymptotically consistent.

When the null hypothesis is not valid, and the identification condition fails outside the designated parameter space,[17] then the limiting behavior of the statistic generally becomes more complicated, although consistency may be guaranteed under conditions that take care the behavior of accumulation points. For instance, if every subsequence of the estimator converges to a non-stchastic parameter value inside Θ,[18] then every subsequence of the statistic diverges to infinity, rendering the test consistent.

2 A Glance at Empirical Likelihood

The example about the Market Entry Game that we have previously encountered, was not explicitly addressed. In this example, the statistical model involves relations defined through integrals-moments, but unlike other examples, these relations are partially inequalities. Generally, such

[17]Something that can be perceived to include as a special case the $\Theta^c = \emptyset$ one.

[18]that may correspond to a parameter value in the alternative hypothesis that satisfies the population moment conditions.

a model can be described by systems of inequality-equality equations of the form $\mathbb{E}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1}$, $\mathbb{E}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2}$. The function m_{eq} is defined as $\mathbf{R}^k \times \Theta \to \mathbb{R}^{q_1}$ and corresponds to the part of the vector-valued function that satisfies the equality moment conditions. Similarly, m_{ineq} is defined as $\mathbf{R}^k \times \Theta \to \mathbb{R}^{q_2}$ and corresponds to the part of the vector-valued function satisfying the inequality moment conditions; the vector inequality is interpreted component-wise, meaning that $\mathbb{E}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2}$ implies every component of $\mathbb{E}(m_{\text{ineq}}(z,\theta))$ is non-negative. Clearly, $m = (m_{\text{eq}}, m_{\text{ineq}})$ and $q_1 + q_2 = q$.[19] Thus, in this Market Entry Game Example, we have $m_{\text{eq}}(z,\theta) := (Z_1Z_2 - (1-\theta_1)(1-\theta_2), Z_1(1-Z_2) - (1-\theta_1)\theta_2)$, $m_{\text{ineq}}(z,\theta) := \theta_2 - Z_1(1-Z_2)$, $q_1 = 2$, $q_2 = 1$.

In such cases-commonly encountered in Econometrics-and assuming proper specification, it is reasonable to expect multiple parameter values satisfying the system of equations and inequalities. If this is the case, the conditions for asymptotic identification become vacuous; $\theta = \theta_0 \Rightarrow \mathbb{E}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1}$, $\mathbb{E}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2}$, but the converse is not true: $\mathbb{E}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1}$, $\mathbb{E}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2} \nRightarrow \theta = \theta_0$. Incidentally, deriving the GMME for such a model is feasible, though the estimator may exhibit complex asymptotic behavior. In this context, Assumption 1.C would not hold, even if 1.A-B are satisfied, potentially leading to different stochastic limits for subsequences of θ_n.

In these models, there exists a set $\Theta_0 \subset \Theta$ such that $\theta \in \Theta_0 \Leftrightarrow \mathbb{E}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1}$, $\mathbb{E}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2}$.[20] Since Θ_0 is not a singleton, the statistical model is

[19] The vector notation here is used somewhat loosely; it is relatively accurate when the function values take the form of row vectors. In earlier parts of the chapter, m was presented as if referring to column vectors.

[20] Recall that we are working under Assumption **??**, implying $\Theta_0 \neq \emptyset$.

referred to as *set-identified*. In such scenarios-which generalize the cases encountered thus far-the objectives of statistical inference may become more intricate. For instance, it may be desirable to identify at least one element of Θ_0 to confirm the validity of the assumptions underlying the game that generated the statistical model in the Market Entry Game Example. Alternatively, it may be necessary to test whether a specific $\theta^\star$ lies within Θ_0. More generally, it may be desirable to test whether $\Theta_{\star\star} \subset \Theta_0$, $\Theta_{\star\star} = \Theta_0$, or $\Theta_{\star\star} \cap \Theta_0 \neq \emptyset$.[21]

The resulting question is how to conduct inference on such set-identified models involving inequality constraints on moment integrals. One approach is through the technology of Empirical Likelihood (EL). This section provides a brief overview of this methodology, focusing on its geometric intuition rather than proofs. For this purpose, the following will be useful:

> The empirical distribution, denoted as $\mathbb{P}_n$, of the sample is the discrete uniform distribution over $\mathbb{R}^k$ with support $\{z_i, i = 1, \ldots, n\}$. This description is not entirely precise. In essence, $\mathbb{P}_n$ is a collection of discrete uniform distributions; for each possible realization of the sample, $\mathbb{P}_n$ is the corresponding discrete uniform distribution, supported on the specific values taken by the sample components. Moreover, it is a random element if the relevant measurable structures are considered; it can be understood as a mapping from $\mathbb{R}^{k\times n}$, equipped with its usual Borel algebra, to the set of probability distributions over $\mathbb{R}^k$, equipped with the Borel algebra induced by the topology of weak convergence. However, such details will not concern us here. What matters is that,

[21] Observe that when Θ_0 is a singleton, all these cases are equivalent.

by the definition of the integral, the empirical expressions for the involved moments are essentially integrals with respect to $\mathbb{P}_n$; specifically, $\frac{1}{n}\sum_{i=1}^{n} m(z_i, \theta) = \mathbb{E}_{\mathbb{P}_n}(m(z, \theta))$ (why?).

If $\mathbb{P}$ and $\mathbb{Q}$ are two discrete distributions over $\mathbb{R}^k$ with the same support, then the Kullback-Leibler (KL) divergence of $\mathbb{Q}$ from $\mathbb{P}$ is defined as:

$$D_{\text{KL}}(\mathbb{P}||\mathbb{Q}) := \sum_{i \in \text{supp}} \log\left(\frac{\mathbb{P}(\{i\})}{\mathbb{Q}(\{i\})}\right)\mathbb{P}(\{i\}) = \mathbb{E}_{\mathbb{P}}\left(\log\frac{\mathbb{P}(z)}{\mathbb{Q}(z)}\right).$$

The $D_{\text{KL}}(\mathbb{P}||\mathbb{Q})$ is also called the relative entropy between $\mathbb{P}$ and $\mathbb{Q}$, and belongs to the broader category of functions on collections of probability distributions that encode notions of "distance" and are referred to as statistical divergences. These typically do not have the structure of a metric (e.g., KL is not symmetric and does not satisfy the triangle inequality), yet due to their properties, they represent certain aspects of the geometry/topology of collections of probability distributions.[ab] The D_{KL} is appropriately extended to any pair of distributions over $\mathbb{R}^k$ and yields finite values if and only if $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ (which is equivalent to that if $\mathbb{Q}(A) = 0$ then $\mathbb{P}(A) = 0$). KL satisfies characteristic inequalities with respect to various metrics on collections of probability distributions. For example, it is proven that the total variation distance between $\mathbb{P}$ and $\mathbb{Q}$-that is, the maximum absolute difference in probabilities assigned to Borel sets-is bounded above by a constant times the square root of $D_{\text{KL}}(\mathbb{P}||\mathbb{Q})$. This observation aligns with the earlier remark regarding the representation of aspects of the associated geometries by this divergence.

[a] It does, however, satisfy the property of being positive-definite-why?
[b] The interested reader is referred to [9] for further details.

KL divergence widely appears as a component in methodologies for constructing statistical inference procedures. Under certain conditions, the expected values of likelihood functions in parametric statistical models are represented as KL divergences between the distributions involved in the model and that corresponding to θ_0. This observation motivates the question of whether it is possible to construct "likelihood functions" in semiparametric models using KL. The concept of empirical likelihood arises as a response to this question. Let us attempt to examine it under the lens of the set-identified model defined by the aforementioned equality-inequality moment relations. Suppose we are interested in testing the following hypothesis structure:

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \notin \Theta_0.$$

which is clearly equivalent to:

$$\begin{aligned} H_0 &: \mathbb{E}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1},\ \mathbb{E}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2} \\ H_1 &: \mathbb{E}(m_{\text{eq}}(z,\theta)) \neq \mathbf{0}_{q_1},\ \text{or } \mathbb{E}(m_{\text{ineq}}(z,\theta)) < \mathbf{0}_{q_2}. \end{aligned} \tag{6}$$

Let $\mathcal{P}_n$ denote the collection of discrete distributions over $\mathbb{R}^k$ that are absolutely continuous with respect to $\mathbb{P}_n$. This collection includes all (potentially stochastic) distributions whose support is some subset of the sample components. If $\theta \in \Theta_{\star\star}$, then the empirical likelihood of the aforementioned model computed at Θ, $\mathcal{EL}_n(\theta)$, is defined as the result of the following optimization problem:

$$\mathcal{EL}_n(\theta) := \min_{\mathbb{P} \in \mathcal{P}_n} D_{\text{KL}}(\mathbb{P}_n || \mathbb{P}), \text{subject to the constraints}$$
$$\mathbb{E}_{\mathbb{P}}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1}, \text{and } \mathbb{E}_{\mathbb{P}}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2}.$$

Hence, the value of the likelihood function at θ results from finding the discrete probability distribution for which the equality-inequality constraints are satisfied and which "deviates" the least from the empirical distribution based on the KL divergence.

If the likelihood value is "sufficiently small," the sample value is considered evidence supporting the null hypothesis. For example, if $\mathbb{E}_{\mathbb{P}}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1}$, and $\mathbb{E}_{\mathbb{P}}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2}$ for $\mathbb{P} = \mathbb{P}_n$, the empirical distribution itself supports the null hypothesis, and the likelihood value in this case is zero (Exercise: Prove this.). Inference for the null hypothesis employs a statistic called the Empirical Likelihood Ratio (ELR) statistic, defined as $\mathcal{ELR}_n(\theta) := 2n\mathcal{EL}(\theta)$. Under assumptions such as iid data, existence of sufficient moments of the random elements involved, and covariance matrix consistency of the empirical moments satisfying equality constraints, it is shown that the asymptotic distribution of the statistic under the null hypothesis resembles the distribution of the minimum of a quadratic form arising in GMME asymptotic theory when θ_0 is a boundary point. In this context, H represents the positive part of the Euclidean space whose dimension equals the number of constraints ultimately satisfied as equalities. The challenge is that this dimension is generally unknown, as it may exceed q_1. However, it can be consistently estimated, and the asymptotic rejection region can be approximated using resampling techniques-see, for example, [11]. Given the above, the testing procedure is straightforward: $\mathcal{ELR}_n$ is computed for the given sample value, and the null hypothesis is rejected if the resulting value lies within the approximate rejection region. It is proven that this test is asymptotically accurate and consistent-see [4]. The same work also proves that, under the independence framework, the test is optimal under certain asymptotic properties not previously discussed. These properties pertain to the rate at which probabilities defining the accuracy and consistency characteristics converge to their respective limits. This optimality feature makes the test particularly preferable to alternatives. Moreover, whether extensions of the test to non-iid environments retain this optimality is, to the best of our knowledge, an interesting and unresolved research question.

The above framework extends to other hypothesis structures related to previously mentioned observations on inference in such statistical models. For instance, if we have:

$$\begin{aligned} H_0 &: \ \Theta_{\star\star} \subseteq \Theta_0 \\ H_1 &: \ \exists \theta \in \Theta_{\star\star} : \theta \notin \Theta_0. \end{aligned}$$

which is clearly equivalent to:

$$\begin{aligned} H_0 &: \ \forall \theta \in \Theta_{\star\star}, \mathbb{E}(m_{\text{eq}}(z,\theta)) = \mathbf{0}_{q_1}, \ \mathbb{E}(m_{\text{ineq}}(z,\theta)) \geq \mathbf{0}_{q_2} \\ H_1 &: \ \exists \theta \in \Theta_{\star\star}, \mathbb{E}(m_{\text{eq}}(z,\theta)) \neq \mathbf{0}_{q_1}, \text{ or } \mathbb{E}(m_{\text{ineq}}(z,\theta)) < \mathbf{0}_{q_2}. \end{aligned} \tag{7}$$

then the test statistic $\sup_{\theta \in \Theta_{\star\star}} \mathcal{ELR}_n(\theta)$ can be used. Based on the aforementioned results, it is possible to derive the asymptotic distribution of this statistic under the current null hypothesis and design an asymptotically accurate and consistent test procedure using resampling techniques.

> *Exersice*: What would the corresponding statistic look like for the hypothesis $\Theta_{\star\star} \cap \Theta_0 \neq \emptyset$?

It should be noted that the optimizations involved in computing the aforementioned statistics may not be trivial! For example, complex operational research methodologies, such as those involving biconvex programming problems,[22] may be required for related tests in econometric models of stochastic dominance.[23]

Finally, it should not be assumed that empirical likelihood techniques are only applicable in set-identified model frameworks like those described above. They can also be used in contexts such as those of previous sections for point estimation, hypothesis testing, etc. For further details, the interested reader may consult [10] and the references therein.

References

[1] Donald WK Andrews. "Estimation when a parameter is on a boundary". In: *Econometrica* 67.6 (1999), pp. 1341–1383.

[2] Donald WK Andrews. "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". In: *Econometrica: Journal of the Econometric Society* (1991), pp. 817–858.

[22]The interested reader may consult [6] for further details.

[23]Stochastic dominance refers to orderings of probability distributions, that in many cases reflect choices w.r.t. large sets of utilities inside the paradigm of Expected Utility. They are thus significant in Economic Theory and Empirical Economics.

[3] Kimiko O Bowman and LR Shenton. "Estimation: Method of moments". In: *Encyclopedia of statistical sciences* 3 (2004).

[4] Ivan A Canay. "EL inference for partially identified models: Large deviations optimality and bootstrap validity". In: *Journal of Econometrics* 156.2 (2010), pp. 408–425.

[5] John S Chipman and MM Rao. "Projections, generalized inverses, and quadratic forms". In: *Journal of Mathematical Analysis and Applications* 9.1 (1964), pp. 1–11.

[6] Jochen Gorski, Frank Pfeuffer, and Kathrin Klamroth. "Biconvex sets and optimization with biconvex functions: a survey and extensions". In: *Mathematical methods of operations research* 66.3 (2007), pp. 373–407.

[7] Lars Peter Hansen. "Large sample properties of generalized method of moments estimators". In: *Econometrica: Journal of the econometric society* (1982), pp. 1029–1054.

[8] Lars Peter Hansen, John Heaton, and Amir Yaron. "Finite-sample properties of some alternative GMM estimators". In: *Journal of Business & Economic Statistics* 14.3 (1996), pp. 262–280.

[9] George G Judge and Ron C Mittelhammer. *An information theoretic approach to econometrics*. Cambridge University Press, 2011.

[10] Yuichi Kitamura. "Empirical likelihood methods in econometrics: Theory and practice". In: (2006).

[11] Dimitris N Politis, Joseph P Romano, and Michael Wolf. *Subsampling*. Springer Science & Business Media, 1999.

[12] Madan L Puri, Carl T Russell, and Thomas Mathew. “Convergence of generalized inverses with applications to asymptotic hypothesis testing”. In: *Sankhyā: The Indian Journal of Statistics, Series A* (1984), pp. 277–286.