

# GMM Estimation

## Contents

- 12.1. Introduction
- 12.2. Formulation
- 12.3. Central limit approximation
- 12.4. Mean value approximation
- 12.5. Approximate inference for testing
- 12.6. GMM Efficiency Bound
- 12.7. Statistical tests based on efficient GMM estimators
- 12.8. An alternative approach to misspecification

 [Download PDF here](#)

---

## 12.1. Introduction

Generalized method moments (GMM) estimation studies a family of estimators constructed from partially specified or partially misspecified models. Since direct application of likelihood methods sometimes can be challenging to construct, GMM methods may be tractable alternatives. It takes a different starting point than a parameterized likelihood function, but is structured to allow for the simultaneous study of a family of estimators. By studying an entire family, we are able to make relative accuracy comparisons among the entire family.

This chapter takes statistical consistency as given. Supporting arguments for this chapter can be obtained with direct extensions of the Law of Large Numbers as described in [Chapter 1 : Laws of Large Numbers and Stochastic Processes](#). Such extensions often entail Laws of Large numbers applied to so-called random functions ( function-valued processes expressed a parameter vector of interest) instead of a random vector.

Throughout this chapter, we will condition on invariant events even though we will suppress this dependence when we write conditional expectations. Given the partially specified or misspecified nature of the model, much more than a simple parameter vector is reflected by this conditioning.<sup>[\[1\]](#)</sup>

## 12.2. Formulation

We study a family of GMM estimators of an unknown parameter vector  $\beta$  constructed from theoretical restrictions on conditional or unconditional moments of functions  $\phi$  that depend on  $\beta$  and on a random vector  $X_t$  that is observable to an econometrician.

As a starting point, we consider a class of restrictions large enough to include examples of both conditional and unconditional moment restrictions. Members of this class take the form

$$E [A_t' \phi(X_t, b)] = 0 \text{ if and only if } b = \beta \quad (12.1)$$

for all sequences of selection matrices  $A \in \mathcal{A}$  where  $A = \{A_t : t \geq 1\}$  and where

- the vector of functions  $\phi$  is  $r$  dimensional.
- the unknown parameter vector  $\beta$  is  $k$  dimensional and in a parameter space  $\mathcal{P}$ .
- $A_t$  denotes a time  $t$  selection matrix for a subset of the valid moment restrictions that is used to construct a particular statistical estimator  $b$  of  $\beta$ .
- $\mathcal{A}$  is a collection of sequences of (possibly random) selection matrices that characterize valid moment restrictions.
- the mathematical expectation is taken with respect to a statistical model that generates the  $\{X_t : t \geq 1\}$  process (captured implicitly by conditioning on invariant events).

A sample counterpart of the population moment conditions [\(12.1\)](#) is

$$\frac{1}{N} \sum_{t=1}^N A_t' \phi(X_t, b_N) = 0. \quad (12.2)$$

Applying a Law of Large Numbers to [\(12.2\)](#) motivates a **generalized method of moments** estimator  $b_N$  of the  $k \times 1$  vector  $\beta$ .

Different sequences of selection matrices  $\{A_t : t \geq 1\}$  and  $\{\tilde{A}_t : t \geq 1\}$  generally give rise to different properties for the estimator  $b_N$ . An exception is when

$$\tilde{A}_t = A_t \mathbb{L}$$

for some  $k \times k$  nonsingular matrix  $\mathbb{L}$ . In this latter case, the same moment conditions are used for estimation and hence will give rise to the same GMM estimator  $b_N$ .

We study limiting properties of estimator  $b_N$  conditioned on a statistical model. In many settings, the parameter vector  $\beta$  only incompletely characterizes the statistical model. In such settings, we are led in effect to implement a version of what is known as **semi-parametric** estimation: while  $\beta$  is the finite-dimensional parameter vector that we want to estimate, we acknowledge that, in addition to  $\beta$ , a potentially infinite-dimensional nuisance parameter vector pins down the complete statistical model on which we condition when we apply the law of large numbers and other limit theorems.

### Example 12.1

#### Unconditional moment restrictions

Suppose that

$$E[\phi(X_t, \beta)] = 0$$

where  $r \geq k$ . Let  $\mathcal{A}_t$  be the set of all constant  $r \times k$  matrices  $\mathbb{A}$  of constants. Rewrite the restrictions as:

$$\mathbb{A}' E[\phi(X_t, \beta)] = 0$$

for all  $r \times k$  matrices  $\mathbb{A}$ . [[Sargan, 1958](#)] and [[Hansen, 1982](#)] assumed moment restrictions like these. For instance,

$$\phi(X_t, \beta) = Z_t \eta(Y_t, \beta)'$$

where  $Z_t$  is an  $r$  dimensional vector of **instrumental variables** and  $\eta(Y_t, \beta)$  is a scalar disturbance term in an equation of interest. The vector of **instrumental variables** are presumed to be uncorrelated with  $Z_t$ , which gives rise to vector of moment conditions. When there are more instrumental variables than parameters, we are led to study a family of estimators rather than a single one.

### Remark 12.1

“Moment matching” estimators are another special case of [Example 12.1](#), an approach that has or at least should have close ties to calibration as is done often in economic dynamics. See [[Hansen and Heckman, 1996](#)] for a discussion of the merits of this link.

Suppose that

$$\phi(X_t, b) = \psi(X_t) - \bar{\psi}(b)$$

where

$$E[\psi(X_t)] = \bar{\psi}(\beta).$$

The random vector  $E[\psi(X_t)]$  defines moments to be matched and  $\bar{\psi}(\beta)$  are population values of those moments under a statistical model with parameter vector  $\beta$ . Often that statistical model is a “structural” economic model with nonlinearities and other complications that, for a given,  $\beta$  make it challenging to compute the moments  $E[\psi(X_t)]$  analytically. To proceed, the proposal is to approximate those moments for a given  $b$  by computing a sample mean from a long simulation of the statistical model at parameter vector  $b$ . By running simulations and computing associated sample means for many alternative  $b$  vectors, we can assemble an approximation to the function  $\bar{\psi}(b)$ . [Lee and Ingram, 1991] and [Duffie and Singleton, 1993] used versions of this approach. Notice that in contrast to some other applications of GMM estimation that allow the appearance of unknown nuisance parameters in the statistical model assumed to generate the data, this approach assumes that, given  $b$ , the model completely determines a sample path that we can at least simulate. This method is used in macroeconomics, corporate finance, and asset pricing, sometimes formally and sometimes informally.

### Example 12.2

#### Conditional moment restrictions

Assume the conditional moment restrictions

$$E[\phi(X_t, \beta) \mid \mathfrak{A}_{t-\ell}] = 0$$

for a particular  $\ell \geq 1$  and  $Y_t = X_t$ . Let  $\mathcal{A}_t$  be the set of all  $r \times k$  matrices,  $A_t$ , of bounded random variables that are  $\mathfrak{A}_{t-\ell}$  measurable. Then the preceding conditional moment restrictions are mathematically equivalent to the unconditional moment restrictions

$$E[A_t' \phi(Y_t, \beta)] = 0$$

for all random matrix processes  $\{A_t : t \geq 0\} \in \mathcal{A}$ . This formulation is due to [Hansen, 1985]. Also see a closely related analysis of [Chamberlain, 1987] with a formal link to semiparametric efficiency bounds.

### Remark 12.2

A common way to construct conditional moment conditions in macro-asset pricing is to construct an  $\ell$  period scalar “stochastic discount factor” as a function of data and an unknown parameter

vector. For instance, see [[Hansen and Singleton, 1982](#)] and [[Hansen and Richard, 1987](#)] for initial applications of this formulation, and more extensive discussions, see the books: [[Cochrane, 2001](#)] and [[Singleton, 2006](#)]. A stochastic discount factor discounts the future in a state-dependent way to capture compensations for exposures to uncertainty. Denote this discount factor by  $\psi(X_t, \beta)$ . This stochastic discount factor may be used to represent asset prices:

$$E[R_t \psi(X_t, \beta) \mid \mathfrak{A}_{t-\ell}] = \mathbf{1}_r$$

where  $R_t$  is an  $r$ -dimensional vector of  $\ell$ -period gross returns and  $\mathbf{1}_r$  is an  $r$  dimensional vector of ones. Let

$$\phi(X_t, b) = R_t \psi(X_t, b) - \mathbf{1}_r.$$

Collections  $\mathcal{A}$  of selection processes for both of these examples satisfy the following “linearity” restriction.

**Restriction 12.1.** If  $A^1$  and  $A^2$  are both in  $\mathcal{A}$  and  $\mathbb{L}_1$  and  $\mathbb{L}_2$  are  $k \times k$  matrices of real numbers, then  $A^1 \mathbb{L}_1 + A^2 \mathbb{L}_2$  is in  $\mathcal{A}$ .

A common practice is to use the approach provided in [Example 12.2](#) while substantially restricting the set of moment conditions used for parameter estimation. One possibility is to create unconditional moment restrictions like those in [Example 12.1](#) from a collection of conditional moment restrictions, and thereby reduce the class of GMM estimators under consideration. For instance, let  $A_t^1$  and  $A_t^2$  be two *ad hoc* choices of selection matrices. Form

$$\phi^+(X_t, b) = \begin{bmatrix} A_t^{1'} \\ A_t^{2'} \end{bmatrix} \phi(X_t, b)$$

where  $X_t$  now includes variables used to construct  $A_t^1$  and  $A_t^2$ . We presume that no linear combination of columns of  $A_t^2$  duplicate any columns in  $A_t^1$ . Otherwise, we would omit such columns and adjust  $\phi^+$  accordingly. Let  $r^+ \geq r$  denote the remaining non-redundant columns. We use  $r^+ \times k$  selection matrices  $\mathbb{A}$  to form moment conditions

$$\mathbb{A}' E(\phi^+(X_t, b)) = 0$$

and study an associated family of GMM estimators. This strategy reduces an infinite number of moment conditions to a finite number. There are extensions of this approach. For instance, we could use more than two  $A_t^j$ 's to construct  $\phi^+$ .

## 12.3. Central limit approximation

The process

$$\left\{ \sum_{t=1}^N A_t' \phi(X_t, \beta) : N \geq 1 \right\}.$$

can be verified to have stationary and ergodic increments conditioned on the statistical model. So there exists a [Proposition 3.1](#) decomposition of the process. Provided that

$$E [A_t' \phi(X_t, \beta)] = 0$$

under the statistical model that generates the data, the trend term in the decomposition of [Proposition 3.1](#) is zero, implying that the martingale dominates the behavior of sample averages for large  $N$ . In particular, [Proposition 3.2](#) gives a central limit approximation for

$$\frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, \beta)$$

Let  $A = \{A_t : t \geq 0\}$  and suppose that

$$\sum_{j=0}^{\infty} E [A_{t+j}' \phi(X_{t+j}, \beta) \mid \mathfrak{A}_t]$$

converges in mean square. Define the one-step-ahead forecast error:

$$G_t(A) = \sum_{j=0}^{\infty} E [A_{t+j}' \phi(X_{t+j}, \beta) \mid \mathfrak{A}_t] - \sum_{j=0}^{\infty} E [A_{t+j}' \phi(X_{t+j}, \beta) \mid \mathfrak{A}_{t-1}]$$

Paralleling the construction of the martingale increment in [Proposition 3.2](#),

$$\frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, \beta) \approx \frac{1}{\sqrt{N}} \sum_{t=1}^N G_t(A)$$

where by the approximation sign  $\approx$  we intend to assert that the difference between the right side and left side converges in mean square to zero as  $N \rightarrow \infty$ . Consequently, the covariance matrix in the central limit approximation is



$$E \left[ G_t(A) G_t(A)' \right].$$

Recall [Restriction 12.1](#). For the preceding construction of the martingale increment, it is straightforward to verify that

$$G_t(A^1 \mathbb{L}_1 + A^2 \mathbb{L}_2) = (\mathbb{L}_1)' G_t(A^1) + (\mathbb{L}_2)' G_t(A^2)$$

follows from the linearity of conditional expectations.

#### Example 12.1 (cont.)

Consider again [Example 12.1](#) in which  $A_t = \mathbb{A}$  for all  $t \geq 0$  and

$$G_t(A) = \mathbb{A}' F_t$$

where

$$F_t = \sum_{j=0}^{\infty} E[\phi(X_{t+j}, \beta) \mid \mathfrak{A}_t] - \sum_{j=0}^{\infty} E[\phi(X_{t+j}, \beta) \mid \mathfrak{A}_{t-1}].$$

Define the covariance matrix

$$\mathbb{V} = E(F_t F_t')$$

and note that

$$E \left[ G_t(A) G_t(A)' \right] = \mathbb{A}' \mathbb{V} \mathbb{A}.$$

#### Example 12.2 (cont.)

In [Example 12.2](#)

$$E[\phi(Y_t, \beta) \mid \mathfrak{A}_{t-\ell}] = 0$$

and hence

$$E[A_t' \phi(Y_t, \beta) \mid \mathfrak{A}_{t-\ell}] = 0$$

whenever entries of  $A_t$  are restricted to be  $\mathfrak{A}_{t-\ell}$  measurable.

Consequently

$$E \left[ A_{t+j}' \phi(Y_{t+j}, \beta) \mid \mathfrak{A}_t \right] = 0$$

for  $j \geq \ell$  so that the infinite sums used to construct  $G_t(A)$  simplify to finite sums.

## 12.4. Mean value approximation

Write

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, b_N) &\approx \frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, \beta) + \frac{1}{N} \sum_{t=1}^N A_t' \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right] \sqrt{N}(b_N - \beta) \\ &\approx \frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, \beta) + \nabla(A)' \sqrt{N}(b_N - \beta) \end{aligned}$$

where

$$\nabla(A) \stackrel{\text{def}}{=} E \left( \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right]' A_t \right).$$

Since

$$\frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, b_N) \approx 0,$$

$$\nabla(A)' \sqrt{N}(b_N - \beta) \approx -\frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, \beta).$$

So long as  $\nabla(A)$  is nonsingular,

$$\sqrt{N}(b_N - \beta) \approx -[\nabla(A)']^{-1} \frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, \beta).$$

This approximation underlies an “efficiency bound” for GMM estimation. Notice that the covariance matrix in a central limit approximation is:



$$\mathbf{cov}(A) = [\nabla(A)']^{-1} E [G_t(A)G_t(A)'] [\nabla(A)]^{-1}$$

We want to know how small we can make this matrix by choosing a selection process.

### Example 12.1 (cont.)

Consider again [Example 12.1](#). In this case  $A_t = \mathbb{A}$  for all  $t \geq 0$  and

$$\nabla(A) = \mathbb{D}'\mathbb{A}$$

where

$$\mathbb{D} \stackrel{\text{def}}{=} E \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right]$$

and

$$\mathbf{cov}(\mathbb{A}) = (\mathbb{A}'\mathbb{D})^{-1} \mathbb{A}'\mathbb{V}\mathbb{A}(\mathbb{D}'\mathbb{A})^{-1}$$

## 12.5. Approximate inference for testing

Using an entirely analogous approach, derive limit approximations geared to testing any “over-identifying restrictions.” Let  $B = \{B_t : t \geq 0\}$  be an  $r \times \tilde{k}$  selection process constructed to test the following vector of  $\tilde{k}$  means.

$$E [B_t' \phi(X_t, \beta)] = 0.$$

**Restriction 12.2.** For any  $\tilde{k} \times k$  matrix of real numbers  $\mathbb{K}$ ,  $B\mathbb{K} \in \mathcal{A}$ .

Thus, we can build selection processes for testing equations from the columns of the process  $B$ .

Suppose that

$$\sum_{j=0}^{\infty} E [B_{t+j}' \phi(X_{t+j}, \beta) \mid \mathcal{A}_t]$$

converges in mean square so that we can apply a central limit approximation.

Construct

$$\tilde{\nabla}(B) \stackrel{\text{def}}{=} E \left( \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right]' B_t \right).$$

Since [Restriction 12.2](#) is satisfied, notice that

$$\tilde{\nabla}(B)\mathbb{K} = \nabla(B\mathbb{K})$$

for all  $\tilde{k} \times k$  matrices  $\mathbb{K}$  of real numbers.

By imitating an earlier argument

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{t=1}^N B_t' \phi(X_t, b_N) &\approx \frac{1}{\sqrt{N}} \sum_{t=1}^N B_t' \phi(X_t, \beta) + \tilde{\nabla}(B)' \sqrt{N}(b_N - \beta) \\ &\approx \frac{1}{\sqrt{N}} \sum_{t=1}^N B_t' \phi(X_t, \beta) \\ &\quad - \tilde{\nabla}(B)' \nabla(A)^{-1} \frac{1}{\sqrt{N}} \sum_{t=1}^N A_t' \phi(X_t, \beta) \\ &\approx \frac{1}{\sqrt{N}} \sum_{t=1}^N \left[ B_t' - \tilde{\nabla}(B)' [\nabla(A)']^{-1} A_t' \right] \phi(X_t, \beta) \end{aligned}$$

This formula includes an explicit adjustment for estimation of  $\beta$ . Notice that if  $A_t = B_t$ , then the right side is zero and the limiting distribution is degenerate. This approximation is used to construct tests that account for having used GMM to estimate a parameter vector  $\beta$ .

#### Example 12.1 (cont.)

Consider again unconditional moment restrictions specified in [Example 12.1](#). Let the selection process for testing be constant over time so that  $B_t = \mathbb{B}$ . Then

$$\frac{1}{\sqrt{N}} \sum_{t=1}^N B_t' \phi(X_t, b_N) \approx \frac{1}{\sqrt{N}} \sum_{t=1}^N \left[ \mathbb{B}' - \mathbb{B}' \mathbb{D} (\mathbb{A}' \mathbb{D})^{-1} \mathbb{A}' \right] \phi(X_t, \beta).$$

#### Remark 12.3

Moment-matching is also accompanied by some form of testing. This practice often occurs in the study of dynamic economic models, but it is also prevalent in other disciplines as the following quote illustrates:

Some hydrologists have suggested a two-step calibration scheme in which the available dependent data set is divided into two parts. In the first step, the independent parameters of the model are adjusted to reproduce the first part of the data. Then in the second step the model is run and the results are compared with the second part of the data. In this scheme, the first step is labeled “calibration,” and the second step is labeled “verification”. [Oreskes *et al.*, 1994].

The formulas just presented for testing support the second “verification” step. Interestingly, while not dismissing it, [Oreskes *et al.*, 1994] are quick to remind readers how about the limitations of this form of verification.

## 12.6. GMM Efficiency Bound

Recall

$$\text{cov}(A) = [\nabla(A)']^{-1} E [G_t(A)G_t(A)'] [\nabla(A)]^{-1}$$

We seek a greatest lower bound on the covariance matrix on the right.

1. Suppose that  $[\nabla(A)']^{-1}$  is nonsingular and impose that

$$[\nabla(A)] = \mathbb{I}$$

If not, post multiply  $A$  by a nonsingular matrix  $\mathbb{J}$ . That leaves the GMM estimator unaltered. Thus, we have

$$\text{cov}(A) = E [G_t(A)G_t(A)']$$

subject to  $[\nabla(A)] = \mathbb{I}$

2. Find an  $A^d$  such that for all  $A \in \mathcal{A}$

$$\nabla(A) = E [G_t(A^d)G_t(A)']. \quad (12.3)$$

3. Form

$$A_t^* = A_t^d (E [G_t(A^d)G_t(A^d)'])^{-1}$$

for all  $A \in \mathcal{A}$ , where we think of [\(12.3\)](#) as a set of first-order (necessary and sufficient) conditions for our constrained optimization. Rather than derive them as such, we essentially use “guess and verify” in what follows.

$$G_t(A^*) = \left( E \left[ G_t(A^d) G_t(A^d)' \right] \right)^{-1} G_t(A^d)$$

and

$$E \left[ G_t(A^*) G_t(A)' \right] = \left( E \left[ G_t(A^d) G_t(A^d)' \right] \right)^{-1}$$

provided that  $[\nabla(A)] = \mathbb{I}$ .

4. Therefore,

$$0 \leq E \left( [G_t(A) - G_t(A^*)][G_t(A) - G_t(A^*)]' \right) = \text{cov}(A) - \left( E \left[ G_t(A^d) G_t(A^d)' \right] \right)^{-1}.$$

**Result 12.1.** Given a solution to equation [\(12.3\)](#)

$$\inf_{A \in \mathcal{A}} \text{cov}(A) = \left( E \left[ G_t(A^d) G_t(A^d)' \right] \right)^{-1} \quad (12.4)$$

#### Remark 12.4

In the result [9.1](#) efficiency bound, we might be tempted to think that  $G_t(A^d)$  plays the same role that the “score vector” increment does in maximum likelihood estimation. But because there is a possibly infinite dimensional vector of nuisance parameters here, a better analogy is that  $G_t(A^d)$  acts much like the residual vector in a regression of parameters of interest score increments on nuisance parameter score increments. By undertaking to infer the parameter vector  $\beta$  from conditional or unconditional moment restrictions, we have purposefully pushed all nuisance parameters into the background.

#### Remark 12.5

The representation

$$E \left( \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right]' A_t \right) = \nabla(A) = E \left[ G_t(A^d) G_t(A)' \right]$$

used to compute the efficiency bound is an application of the Riesz Representation Theorem. To understand this, introduce the  $k$ -dimensional coordinate vectors  $\mathbf{u}_i$  for  $i = 1, 2, \dots, k$  and consider:

$$(\mathbf{u}_i)' \nabla(A) \mathbf{u}_j = E \left( \left[ \frac{\partial \phi}{\partial b_i}(X_t, \beta) \right] \cdot A_t \mathbf{u}_j \right). \quad (12.5)$$

Note that

- The integer  $i$  selects the coordinate of  $b$  with respect to which we are differentiating.
- If  $A^1$  and  $A^2$  are both in  $\mathcal{A}$ , then so are linear combinations. Therefore  $(\mathbf{u}_i)' \nabla(A) \mathbf{u}_j$  is a linear functional defined on a linear space of random variables of the form  $\phi(X_t, \beta)' A_t \mathbf{u}_j$  for a given  $i$ .
- The martingale approximations for the scalar process with stationary increments

$$\left\{ \sum_{t=1}^N \phi(X_t, \beta)' A_t \mathbf{u}_j : N \geq 1 \right\}$$

has martingale increment  $G_t(A) \mathbf{u}_j$ .

- The Riesz Representation Theorem asserts that the linear functional  $(\mathbf{u}_i)' \nabla(A) \mathbf{u}_j$  can be represented as an inner product

$$(\mathbf{u}_i)' \nabla(A) \mathbf{u}_j = E [R_t G_t(A) \mathbf{u}_j]$$

where the scalar random variable  $R_t$  is in the mean square closure of

$$\{G_t(A) \mathbf{u}_j : A \in \mathcal{A}\}.$$

- We can represent  $R_t$  as

$$R_t = G_t(A^d) \mathbf{u}_j$$

for some selection process  $A_d \in \mathcal{A}$  or more generally as a limit point of a sequence of such selection processes.

The preceding construction pins down row  $j$  of  $A^d$ . Repeating an analogous construction for each  $j = 1, 2, \dots, k$  gives the selection matrix  $A^d$ .

The GMM efficiency bound presumed that we could solve equation [\(12.5\)](#). The Riesz Representation Theorem requires that  $R_t$  be in a mean square closure of a linear space. Provided that the linear functionals  $(\mathbf{u}_i)' \nabla(A) \mathbf{u}_j$  are mean square continuous, the efficiency bound can be

represented in terms of the limit point of a sequence of GMM estimators associated with alternative selection processes even when the limit points are not attained.

### Example 12.1 (cont.)

Consider [Example 12.1](#) in which we assumed that  $A_t = \mathbb{A}$ . Then

$$\mathbb{A}'\mathbb{V}\mathbb{A}^d = \mathbb{A}'\mathbb{D}.$$

Therefore,

$$\mathbb{A}^d = \mathbb{V}^{-1}\mathbb{D}$$

and the GMM efficiency bound is

$$(\mathbb{D}'\mathbb{V}^{-1}\mathbb{D})^{-1}.$$

### Example 12.2 (cont.)

Consider again [Example 12.2](#) in the special case in which  $\ell = 1$  so that the conditional moment condition of interest is:

$$E[\phi(X_t, \beta) \mid \mathfrak{A}_{t-1}] = 0.$$

Let

$$E[\phi(X_t, \beta)\phi(X_t, \beta)' \mid \mathfrak{A}_{t-1}] = V_{t-1},$$

which we assume to be nonsingular. To compute the efficiency bound, we wish to solve the following equation for  $A_t^d$

$$E\left(A_t^{d'} V_{t-1} A_t\right) = \nabla(A) = E\left(\left[\frac{\partial \phi}{\partial b'}(X_t, \beta)\right]' A_t\right). \quad (12.6)$$

Given the flexibility in the choice of the random  $A_t$  with entries that are  $\mathfrak{A}_{t-1}$  measurable, this equation is equivalent to

$$V_{t-1}A_t^d = E \left( \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right] \mid \mathfrak{A}_{t-1} \right)$$

where we have taken transposes of the expressions in [\(12.6\)](#). Thus

$$A_t^d = (V_{t-1})^{-1} E \left( \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right] \mid \mathfrak{A}_{t-1} \right)$$

and the efficiency bound is:

$$\left[ E \left( \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right]' \mid \mathfrak{A}_{t-1} \right) (V_{t-1})^{-1} E \left( \left[ \frac{\partial \phi}{\partial b'}(X_t, \beta) \right] \mid \mathfrak{A}_{t-1} \right) \right]^{-1}.$$

Next consider a conditional moment justification for **two-stage least squares**. Add the following special restrictions. Suppose that  $r = 1$  and that  $V_{t-1} = \mathbf{v} > 0$  where  $\mathbf{v}$  is constant. Further suppose that

$$\phi(X_t, b) = Y_t^1 - Y_t^2 \cdot b$$

Finally, suppose that

$$E(Y_t^2 \mid \mathfrak{A}_{t-1}) = \Pi Z_{t-1}$$

where  $Z_{t-1}$  has more entries than  $Y_t^2$ . Notice that  $\Pi$  can be computed as a least squares regression. Then

$$A_t^d = \left( \frac{1}{\mathbf{v}} \right) Z_{t-1}' \Pi'$$

The scaling by  $\frac{1}{\mathbf{v}}$  is inconsequential to the construction of a selection process. The matrix of regression coefficients can be replaced by the finite sample least squares regression coefficients without altering the statistical efficiency.

To obtain this rationale for two-stage least squares, we had to impose a special structure, one that does not prevail in many important applications. For instance, suppose that  $V_{t-1}$  depends on conditioning information so that a form of conditional heteroskedasticity is present. That dependence shows up in essential ways in how  $A_t^d$  should be constructed. Further, suppose that the expectation  $E[Y_t^2 \mid \mathfrak{A}_{t-1}]$  depends nonlinearly on  $Z_{t-1}$ . In that case, to attain or to approximate the efficiency bound, a least squares regression should account for potential nonlinearity.



### Remark 12.6

To implement the conditional moment version of the GMM efficiency bound requires the estimation of conditional moments for which the model may not provide functional forms. Reliable nonparametric estimation becomes particularly challenging in high dimensions. Suppose instead that a researcher adopts a convenient parametric approximation to these conditional moments. While this will induced a form of misspecification, it will not necessarily undermine either the consistency of the resulting GMM estimator or its asymptotic distribution. The misspecification may only make the resulting estimator only cause a reduction in the statistical efficiency of the estimator as measured by the asymptotic covariance matrix of the resulting GMM estimator.

### Example 12.1 (cont.)

Suppose now that  $\ell = 2$  and consider an unconditional moment formulation.

Then even if the covariance structure is homoskedastic and conditional expectations are linear, the two-squares least square approach will no longer be statistically efficient. We illustrate why by mapping into the framework of [Example 12.1](#).

Use as our  $\phi$  function

$$\phi(X_t, b) = Z_{t-2} \left[ Y_t^1 - (Y_t^2)' b \right]$$

in forming an unconditional moment restriction and implementing constructing the efficient GMM estimator where the entries of  $Z_{t-2}$  and in the date  $t - 2$  conditioning information set ( $\mathcal{A}_{t-2}$  measurable). Now form:

$$\mathbb{V} = E \left[ \phi(X_t, \beta) \phi(X_t, \beta)' \right] + E \left[ \phi(X_t, \beta) \phi(X_{t+1}, \beta)' \right] + E \left[ \phi(X_t, \beta) \phi(X_{t-1}, \beta)' \right]$$

pertinent to the central limit approximation. Then an efficient selection matrix  $\mathbb{A}^d$  is given by:

$$\mathbb{A}^d = \mathbb{V}^{-1} E \left[ Z_{t-2} (Y_t^2)' \right]$$

The matrix  $\mathbb{A}^d$  is typically **not** proportional to a vector of regression coefficients of  $Y_t^2$  on  $Z_{t-2}$ , as presumed in the justification for two-stage least squares. The temporal dependence removes this connection. Hence, a commonly-used measure of “instrument relevance” obtained by regressing the endogenous  $Y_t^2$  onto  $Z_{t-2}$  is no longer a valid diagnostic .

A special case of this analysis is when  $Y_t^2$  in the date  $t - 2$  conditioning information set. One estimator could be constructed by setting  $Z_{t-2} = Y_t^2$ . The standard two-stage least squares estimator now is simply the ordinary least squares estimator. Next expand  $Z_{t-2}$  to be:

$$Z_{t-2} = \begin{bmatrix} Y_t^2 \\ Y_{t-1}^2 \end{bmatrix}.$$

The  $\mathbb{A}^d$  selection matrix will typically include  $Y_{t-1}^2$  in the estimation in spite of the fact that this variable is not needed in an initial least squares regression of  $Y_t^2$  onto  $Z_{t-2}$ .

Since there is great flexibility in the construction of  $Z_{t-\ell}$ , there is further scope for efficiency gains attained by using the conditional moment formulation of the family of GMM estimators. [\[Hansen and Singleton, 1996\]](#) construct the efficiency bound in [Example 12.2](#) for a linear data generating process.

## 12.7. Statistical tests based on efficient GMM estimators

First suppose that we have statistically efficient selection process. Thus the selection matrix is some nonsingular matrix transformation of  $A^d$  where  $A^d$  satisfies the “first-order conditions” [\(12.3\)](#). Recall the approximation

$$\frac{1}{\sqrt{N}} \sum_{t=1}^N B_t' \phi(X_t, b_N) \approx \frac{1}{\sqrt{N}} \sum_{t=1}^N \left[ B_t' - \tilde{\nabla}(B)' [\nabla(A^d)']^{-1} A_t^{d'} \right] \phi(X_t, \beta),$$

which includes an adjustment for estimation. Let  $\tilde{G}_t(B)$  denote the increment in the martingale approximation for

$$\sum_{t=1}^N B_t' \phi(X_t, \beta).$$

We use the representation implied by [\(12.3\)](#) to write:

$$\tilde{\nabla}(B) = E [G_t(A^d) G_t(B)'].$$

These constructions allow us to write:

$$\frac{1}{\sqrt{N}} \sum_{t=1}^N B_t' \phi(X_t, b_N) \approx \frac{1}{\sqrt{N}} \sum_{t=1}^N \hat{G}_t(B) \tag{12.7}$$

where

$$\widehat{G}_t(B) \stackrel{\text{def}}{=} \widetilde{G}_t(B) - E \left[ \widetilde{G}_t(B) G_t(A^d)' \right] \left( E \left[ G_t(A^d) G_t(A^d)' \right] \right)^{-1} G_t(A^d)$$

The term,  $\widehat{G}_t(B)$ , that appears inside the sum on the right side of (12.7) is the population least squares residual from regressing  $\widetilde{G}_t(B)$  onto  $G_t(A^d)$ . This regression residual can also be interpreted as a martingale increment for a stationary increments process.

Suppose that  $\widehat{G}_t(B)$  has a nonsingular covariance matrix. Consider the quadratic form used for building a test:

$$\frac{1}{N} \left[ \sum_{t=1}^N \phi(X_t, b_N)' B_t \right] \left( E \left[ \widehat{G}_t(B) \widehat{G}_t(B)' \right] \right)^{-1} \left[ \sum_{t=1}^N B_t' \phi(X_t, b_N) \right] \Rightarrow \chi^2(\tilde{k}).$$

This test can be implemented in practice by replacing  $E \left[ \widehat{G}_t(B) \widehat{G}_t(B)' \right]$  with a statistically consistent estimator of it. There is an equivalent way to represent this quadratic form:

$$\frac{1}{N} \sum_{t=1}^N \phi(X_t, b_N)' \begin{bmatrix} B_t & A_t^d \end{bmatrix} \left[ E \left( \begin{bmatrix} \widetilde{G}_t(B) \\ G_t(A^d) \end{bmatrix} \begin{bmatrix} \widetilde{G}_t(B)' & G_t(A^d)' \end{bmatrix} \right) \right]^{-1} \left[ \sum_{t=1}^N \begin{bmatrix} B_t' \\ A_t^{d'} \end{bmatrix} \phi(X_t, b_N) \right]$$

This equivalence follows because the inverse of the covariance matrix for the regression error  $\widehat{G}_t(B)$  is the upper diagonal block of the inverse of the covariance matrix:

$$E \left( \begin{bmatrix} \widetilde{G}_t(B) \\ G_t(A^d) \end{bmatrix} \begin{bmatrix} \widetilde{G}_t(B)' & G_t(A^d)' \end{bmatrix} \right)$$

### Example 12.1 (cont.)

Consider [Example 12.1](#) again. We have already shown that

$$\mathbb{A}^d = \mathbb{V}^{-1} \mathbb{D}.$$

Suppose that we choose  $\mathbb{B}$  with dimension  $r \times (r - k)$  so that

$$[\mathbb{A}^d \quad \mathbb{B}]$$

has full rank. Then

$$\frac{1}{N} \sum_{t=1}^N \phi(X_t, b_N)' \mathbb{V}^{-1} \sum_{t=1}^N \phi(X_t, b_N)' \Rightarrow \chi^2(r - k).$$

If we replace  $b_N$  with  $\beta$  on the left side of the above limit we find

$$\frac{1}{N} \sum_{t=1}^N \phi(X_t, \beta)' \mathbb{V}^{-1} \sum_{t=1}^N \phi(X_t, \beta)' \Rightarrow \chi^2(r)$$

The difference in the resulting  $\chi^2$  distribution emerges because estimating  $k$  free parameters reduces degrees of freedom by  $k$ . It is straightforward to show that

$$\frac{1}{N} \sum_{t=1}^N \phi(X_t, \beta)' \mathbb{V}^{-1} \sum_{t=1}^N \phi(X_t, \beta)' - \frac{1}{N} \sum_{t=1}^N \phi(X_t, b_N)' \mathbb{V}^{-1} \sum_{t=1}^N \phi(X_t, b_N)' \Rightarrow \chi^2(k),$$

an approximation that is useful for constructing confidence sets for GMM estimates of parameter vector  $\beta$ .

#### Remark 12.7

To continue our study of [Example 12.1](#), form the population problem:

$$\min_b E[\phi(X_t, b)]' \mathbb{V}^{-1} E[\phi(X_t, b)].$$

This has a minimizer at  $b = \beta$  provided that the unconditional moment conditions are satisfied. If  $b = \beta$  is the only possible parameter vector that satisfies the population moment conditions, then  $b = \beta$  is the unique solution to the population minimization problem stated here. Suppose that we construct an estimator by solving a minimization problem:

$$\min_b \frac{1}{N} \sum_{t=1}^N \phi(X_t, b)' \mathbb{V}^{-1} \sum_{t=1}^N \phi(X_t, b). \quad (12.8)$$

First-order necessary conditions are

$$\frac{1}{N} \sum_{t=1}^N \left[ \frac{\partial \phi}{\partial b'}(X_t, b_N) \right]' \mathbb{V}^{-1} \sum_{t=1}^N \phi(X_t, b_N) = 0.$$

Assume that we already know that the solution  $b_N$  of the above first-order conditions provides a consistent estimator of parameter vector  $\beta$ . Then we can show that

$$\frac{1}{N} \sum_{t=1}^N \left[ \frac{\partial \phi}{\partial b'}(X, b_N) \right] \rightarrow \mathbb{D}$$

where convergence is with probability one. Thus, in this case the implied selection matrix

$$\mathbb{A} = \mathbb{V}^{-1} \mathbb{D}$$

provides an estimator that attains the efficiency bound. The limiting distribution of the minimizer of criterion (12.8) is  $\chi^2$  with  $r - k$  degrees of freedom. Thus this gives a combined approach to estimation and testing

#### Remark 12.8

There is an interesting variation of the approach described in [Remark 12.7](#). For any  $b$ , let  $\mathbb{V}(b)$  be the population covariance matrix in the martingale increment used in the Central Limit approximation for the process

$$\frac{1}{\sqrt{N}} \sum_{t=1}^N \phi(X_t, \beta).$$

Assume that  $\mathbb{V}(b)$  is nonsingular for every  $b$  in a parameter space. Form the population minimization problem:

$$\min_b E[\phi(X_t, b)]' [\mathbb{V}(b)]^{-1} E[\phi(X_t, b)].$$

If  $b = \beta$  is the only vector that satisfies the associated population first-order conditions, then  $b = \beta$  is again the unique solution to the above population minimization problem.

Now form sample counterparts of both  $E[\phi(X_t, b)]$  and  $\mathbb{V}(b)$  as functions of  $b$ . Minimizing a sample counterpart of the above population minimization problem gives rise to a “continuously-updated GMM estimator”. See [[Hansen et al., 1996](#)]. The parameter vector and an appropriately scaled minimized objective function have the same limiting distributions as those described in [Remark 12.7](#).

#### Remark 12.9

It can be numerically challenging to find minimizers to the continuously-updated GMM estimator in high dimensions. Similarly, it can be difficult to construct confidence sets based on this objective. As an [\[Chernozhukov and Hong, 2003\]](#) and [\[Chen et al., 2018\]](#) devise and justify simulation-based methods for inference applicable for the continuously-weighted GMM objective function. They do so by adapting insights from simulation-based approaches for Bayesian inferences. The Bayesian-type calculations can be numerically more tractable [\[Chernozhukov and Hong, 2003\]](#) treat the continuously-updated objective function as a “log-likelihood function”, and then use large sample approximations to justify the Bayesian-like calculations. They justify this approach even though the continuously-updated objective is not formally a log-likelihood function. [\[Chen et al., 2018\]](#) show how to modify this approach to attain additional robustness and reliability. An attractive feature of these methods is that they exploit the shape of the continuously-updated GMM objective function, but in so doing they impose a “prior” distribution to help guide this exploration.

#### Remark 12.10

When applications of GMM methods call for general (but weak) forms of temporal dependence, the reliable estimation of the covariance matrix  $\mathbb{V}$  needed for central limit approximation can be very challenging. Various researchers have proposed methods that come from what is called “spectral density” estimation. Within this latter literature, the matrix  $\mathbb{V}$  is the spectral density matrix for the process  $\{\phi(X_t, \beta) : t \geq 0\}$ . Spectral methods are local in nature and can be notoriously unreliable for high dimensions (large values of  $r$ ). Some form of time-series approximation may be necessary for such estimation environments.

In the case of moment matching, assuming a correct specification, one could use very large sample simulations conditioned on each of the hypothetical parameters to approximate the construction of  $\mathbb{V}(b)$ . Remarkably, this is not often done in economics, even though it could improve the quality of the inferences.

## 12.8. An alternative approach to misspecification

We modify the target of estimation to include an estimate of the underlying probability distribution. At the same time as we estimate the parameter vector, we estimate the distribution of the observable data that supports this estimation. We relax the assumption that the moment conditions are satisfied under the data generating process and instead find distributions that are potentially statistically close to the data generating process. In many models of interest, the expectations used for the moment conditions could be the subjective beliefs of the agents “inside the model” with expectations that differ from the actual data generating process.

There is a substantial literature on what is called generalized empirical likelihood that addresses this, with the primary motivation to improve second-order efficiency, and constructs that we do not study here. Since



this literature features a finite number of unconditional moment condition, these second-order efficiency gains are applied to a second-best formulation of statistical efficiency.

Instead of focusing on a refinements inferences, we following a suggestion in [\[Hansen, 2014\]](#) by thinking of the beliefs as those of economic agents within the models we build. Thus we use special cases of the generalized empirical likelihood methods as diagnostics that are informative about potential misspecification. With this aim in mind, relative entropy divergence provides one convenient way to proceed. We also make a revealing comparison to outcomes when we use a quadratic-divergence measure.

## 12.8.1. Relative entropy divergence

Consider the following population alternative proposed by [\[Kitamura and Stutzer, 1997\]](#) to the GMM approach we described so far:

$$\begin{aligned} \min_{b \in \mathcal{P}} \min_{M \geq 0} & E[M(\log M)] \\ & E[M\phi(X_t, b)] = 0 \\ & E(M - 1) = 0. \end{aligned}$$

Given multipliers,  $\lambda$ , and  $\zeta$  on the constraints, the first-order conditions for the inner minimization are:

$$\lambda \cdot \phi(X_t, b) + \zeta + \log(M) + 1 = 0.$$

Solving for the minimizing  $M(b, \lambda, \zeta)$  given the multipliers  $\lambda$  and  $\zeta$  gives:

$$M(b, \lambda, \zeta) = \exp[-\lambda \cdot \phi(X_t, b)] \exp(-\zeta - 1)$$

Imposing the  $EM = 1$  constraint implies that

$$M(b, \lambda) = \frac{\exp[-\lambda \cdot \phi(X_t, b)]}{E(\exp[-\lambda \cdot \phi(X_t, b)])}$$

with an inner minimization problem replaced by

$$\max_{\lambda \in \mathbb{R}^r} -\log E(\exp[-\lambda \cdot \phi(X_t, b)]).$$

Thus the counterpart to a GMM estimator solves the min-max problem:

$$\min_{b \in \mathcal{P}} \max_{\lambda \in \mathbb{R}^r} -\log E(\exp[-\lambda \cdot \phi(X_t, b)]).$$



[[Kitamura and Stutzer, 1997](#)] establish that the minimizing solution to a sample counterpart to this population objective has the same first-order properties as an efficient GMM estimator when  $\{\phi(X_t, \beta) : t \geq 0\}$  is itself a martingale difference sequence.

## 12.8.2. Quadratic divergence

Consider a counterpart with a quadratic divergence to the problem that we just analyzed:

$$\begin{aligned} \min_{b \in \mathcal{P}} \min_{M \geq 0} & \frac{1}{2} E[(M - 1)^2] \\ & E[M\phi(X_t, b)] = 0 \\ & E(M - 1) = 0. \end{aligned}$$

Given multipliers,  $\lambda$ , and  $\zeta$  on the constraints, the first-order conditions for the inner minimization are:

$$\lambda \cdot \phi(X_t, b) + \zeta + M - 1 = 0.$$

Thus

$$M(b, \lambda, \zeta) = 1 - \zeta - \lambda \cdot \phi(X_t, b). \tag{12.9}$$

Since  $M$  has mean one,

$$\zeta = -\lambda \cdot E\phi(X_t, b),$$

and thus

$$M(b, \lambda) - 1 = -\lambda \cdot [\phi(X_t, b) - E\phi(X_t, b)].$$

Let

$$\begin{aligned} \mathbb{V}(b) &\stackrel{\text{def}}{=} E([\phi(X_t, b) - E[\phi(X_t, b)]]\phi(X_t, b)') \\ &= E([\phi(X_t, b) - E[\phi(X_t, b)]][\phi(X_t, b) - E[\phi(X_t, b)]]'), \end{aligned}$$

which is the covariance matrix of the random vector:  $\phi(X_t, b)$ . We take the moment conditions not to be redundant so that  $\mathbb{V}(b)$  is nonsingular. Imposing the constraint that under the probability measure induced by  $M(b, \lambda)$ ,  $\phi(X_t, b)$ , has mean zero gives the formula;

$$\mathbb{V}(b)\lambda = E[\phi(X_t, b)].$$

Thus

$$\begin{aligned} M(b, \lambda) - 1 &= -E[\phi(X_t, b)]' \mathbb{V}(b)^{-1} [\phi(X_t, b) - E\phi(X_t, b)] \\ \frac{1}{2} [(M - 1)^2] &= \frac{1}{2} E[\phi(X_t, b)]' \mathbb{V}(b)^{-1} E[\phi(X_t, b)], \end{aligned} \quad (12.10)$$

and the  $b$  of interest is determined by:

$$\min_{b \in \mathcal{P}} \frac{1}{2} E[\phi(X_t, b)]' \mathbb{V}(b)^{-1} E[\phi(X_t, b)].$$

This is recognizable as the population counterpart to the continuously-updated objective for GMM. For instance, see [\[Hansen et al., 1996\]](#).

So far, we have ignored the restriction that  $M \geq 0$ . The solution for  $M(b, \lambda)$  implied by [\(12.10\)](#) has mean one, but it could be negative with positive probability. Imposing nonnegativity potentially alters the solution for  $M$  and could raise the objective for inferring  $b$ . In this case the first-order conditions [\(12.9\)](#) are modified to be:

$$M(b, \lambda, \zeta) = \begin{cases} 1 - \zeta - \lambda \cdot \phi(X_t, b) & \text{if positive} \\ 0 & \text{otherwise.} \end{cases}$$

We no longer have a quasi-analytical solution, but instead solve:

$$\min_{b \in \mathcal{P}} \max_{\lambda, \zeta} \frac{1}{2} E \left( [M(b, \lambda, \zeta) - 1]^2 \right)$$

### 12.8.3. Using the recovered probabilities as diagnostics for misspecification

[\[Brown and Back, 1993\]](#) suggest using the implied probabilities from a GMM estimation, implemented with a quadratic divergence, as a model diagnostic. Take as a starting point the empirical distribution obtained by assigning  $1/N$  to each of the realized  $X_t$ 's. Then scale these weight the corresponding  $M$  computed at the GMM estimator of  $\beta$  and isolate where the big adjustments are. This can inform an applied researcher as to what aspects of the empirical distribution the model finds particularly challenging.

As alternative following a formulation of [\[Chen et al., 2020\]](#). Consider the relative entropy divergence. A probability distribution is characterized by the expectations it assigns to functions of the underlying random vector, say  $X_t$ . Take the minimum divergence outcome as a starting point and inflate by some percentage. Call this divergence  $\kappa$ . For a rich class of functions,  $\tilde{\phi}$ , and hypothetical parameter vector,  $b$ , solve

$$\begin{aligned} \min_{b \in \mathcal{P}} \min_{M \geq 0} & E \left[ M \tilde{\phi}(X_t, b) \right] \\ & E \left[ M \log M \right] \leq \kappa \\ & E \left[ M \phi(X_t, b) \right] = 0 \\ & E(M - 1) = 0. \end{aligned}$$

This provides a sharp upper bound on  $E[M \tilde{\phi}(X_t, b)]$ . To get a sharp lower bound, repeat the same calculation by computing a sharp upper bound for  $E[M \tilde{\phi}(X_t, b)]$  and multiply the outcome by minus one.

We may leverage our previous calculations by proceeding differently. First solve the minimum divergence problem and compute the implied expectation for  $E[M, \tilde{\phi}(X_t, b)] = \underline{r}$ . Solve:

$$\begin{aligned} \min_{b \in \mathcal{P}} \min_{M \geq 0} & [M(\log M)] \\ & E \left[ M \phi(X_t, b) \right] = 0 \\ & E \left[ M \psi(X_t, b) \right] - r = 0 \\ & E(M - 1) = 0. \end{aligned}$$

for  $r > \underline{r}$ . The added constraint will induce a larger divergence bound. By increasing  $r$  we may attain a divergence of  $\kappa$ . With formulation, we have an immediate extension of our previous analysis leading to the following problem

$$\min_{b \in \mathcal{P}} \max_{\lambda \in \mathbb{R}^r, \tilde{\lambda} \in \mathbb{R}} -\log E \left( \exp \left[ -\lambda \cdot \phi(X_t, b) - \tilde{\lambda} \tilde{\phi}(X_t, b) + \tilde{\lambda} r \right] \right)$$

As a special case, we may set  $\tilde{\phi}(X_t, b)$  equal to one of the coordinates of  $b$  and deduce implied upper and lower bounds on the different parameter coefficients.

[Chen et al., 2024] describe inferential methods that support such an analysis expressed in terms of large sample approximations.

#### Remark 12.11

So far, we have treated this as unconditional problem including both the moment conditions and the divergence measures. [Chen et al., 2020] show how to extend this analysis to the case with conditional moment restrictions along with an intertemporal measure of statistical divergence.

[1] See [Hansen, 2000] and [Hansen, 2008] for overviews of GMM estimation.

