Chapter 9

GENERALIZED METHOD OF MOMENTS

9.1 Asymptotic Properties of GMM Estimators

9.1.1 Moment Restriction and GMM Estimators

To motivate GMM estimation, consider Hansen and Singleton's (1982) Consumption-Based Capital Asset Pricing Model (C-CAPM). A representative agent maximizes

$$\sum_{t=1}^{\infty} \beta^t E(U(c_t)|\mathrm{I}_0) \tag{9.1}$$

subject to a budget constraint. Hansen and Singleton (1982) use an isoelastic intraperiod utility function

$$U(c_t) = \frac{1}{1-\alpha}(c_t^{1-\alpha} - 1), \tag{9.2}$$

where c_t is real consumption at date t, β is a discount factor and $\alpha > 0$ is the reciprocal of the intertemporal elasticity of substitution (α is also the relative risk aversion coefficient for consumption in this model). The standard Euler equation for the optimization problem is

$$\frac{E[\beta c_{t+1}^{-\alpha} R_{t+1}|\ \mathrm{I}_t]}{c_t^{-\alpha}} = 1, \tag{9.3}$$

where R_{t+1} is the gross real return of an asset and I_t is an information set available at time t. This Euler equation can be rearranged as

$$E[\beta(\frac{c_{t+1}}{c_t})^{-\alpha}R_{t+1} - 1|\ \mathrm{I}_t] = 0. \tag{9.4}$$

Let $\mathbf{z}_t$ be a vector of variables whose values are known at time t. Then $\mathbf{z}_t \in \mathrm{I}_t$ and

$$E[\mathbf{z}_t\{\beta(\frac{c_{t+1}}{c_t})^{-\alpha}R_{t+1} - 1\}|\ \mathrm{I}_t] = \mathbf{0}. \tag{9.5}$$

By the law of iterative expectations, we obtain the orthogonality conditions to be used in GMM estimation,

$$E[\mathbf{z}_t\{\beta(\frac{c_{t+1}}{c_t})^{-\alpha}R_{t+1} - 1\}] = \mathbf{0}. \tag{9.6}$$

Let $\{\mathbf{x}_t : t = 1, 2, \cdots\}$ be a stationary and ergodic vector stochastic process, $\mathbf{b}_0$ be a p-dimensional vector of the parameters to be estimated, and $f(\mathbf{x}_t, \mathbf{b})$ a q-dimensional vector of functions. We refer to $\mathbf{u}_t = f(\mathbf{x}_t, \mathbf{b}_0)$ as the disturbance of GMM. Consider the (unconditional) moment restrictions

$$E(f(\mathbf{x}_t, \mathbf{b}_0)) = \mathbf{0}. \tag{9.7}$$

For example, in the Hansen and Singleton (1982) case, $\mathbf{x}_t = (\frac{c_{t+1}}{c_t}, R_{t+1}, \mathbf{z}_t')'$, $\mathbf{b}_0 = (\beta, \alpha)'$, and $f(\mathbf{x}_t, \mathbf{b}_0) = \mathbf{z}_t\{\beta(\frac{c_{t+1}}{c_t})^{-\alpha}R_{t+1} - 1\}$.

Suppose that a law of large numbers can be applied to $f(\mathbf{x}_t, \mathbf{b})$ for all admissible $\mathbf{b}$, so that the sample mean of $f(\mathbf{x}_t, \mathbf{b})$ converges to its population mean:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}) = E(f(\mathbf{x}_t, \mathbf{b})) \tag{9.8}$$

with probability one (or, in other words, almost surely). The basic idea of GMM estimation is to mimic the moment restrictions in (9.7) by minimizing a quadratic form of the sample means

$$J_T(\mathbf{b}) = \{\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b})\}'\mathbf{W}_T\{\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b})\} \tag{9.9}$$

with respect to $\mathbf{b}$, where $\mathbf{W}_T$ is a positive semidefinite matrix that satisfies

$$\lim_{T\to\infty} \mathbf{W}_T = \mathbf{W}_0 \tag{9.10}$$

with probability one for a positive definite matrix $\mathbf{W}_0$. The matrices $\mathbf{W}_T$ and $\mathbf{W}_0$ are both referred to as the distance or weighting matrix. The GMM estimator, $\mathbf{b}_T$, is the solution of the minimization problem in (9.9). Under fairly general regularity conditions, the GMM estimator $\mathbf{b}_T$ is a consistent estimator for arbitrary distance matrices.[1] The selection of the distance matrix which yields an (asymptotically) efficient GMM estimator is discussed below in Section 9.1.3.

9.1.2 Asymptotic Distributions of GMM Estimators

Suppose that a central limit theorem applies to the disturbance of GMM, $\mathbf{u}_t = f(\mathbf{x}_t, \mathbf{b}_0)$, so that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{u}_t$ has an (asymptotic) normal distribution with mean zero and the covariance matrix $\mathbf{\Omega}$ in large samples.[2] If $\mathbf{u}_t$ is serially uncorrelated, $\mathbf{\Omega} = E(\mathbf{u}_t\mathbf{u}_t')$. If $\mathbf{u}_t$ is serially correlated,

$$\mathbf{\Omega} = \lim_{j\to\infty}\sum_{-j}^{j} E(\mathbf{u}_t\mathbf{u}_{t-j}'). \tag{9.11}$$

Some authors refer to $\mathbf{\Omega}$ as the long run covariance matrix of $\mathbf{u}_t$. Let $\mathbf{\Gamma} = E(\frac{\partial f(\mathbf{x}_t,\mathbf{b}_0)}{\partial \mathbf{b}'})$ be the expectation of the $q \times p$ matrix of the derivatives of $f(\mathbf{x}_t, \mathbf{b}_0)$ with respect to $\mathbf{b}$ and assume that $\mathbf{\Gamma}$ has full column rank. Under suitable regularity conditions,

[1] Some regularity conditions that are important for applied researchers will be discussed in Section 9.3

[2] An advantage of the GMM estimation is that a strong distributional assumption such that $\mathbf{u}_t$ is normally distributed is not necessary.

$\sqrt{T}(\mathbf{b}_T - \mathbf{b}_0)$ converges in distribution to a normal distribution with mean zero and the covariance matrix

$$Cov(\mathbf{W}_0) = (\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma})^{-1}(\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Omega}\mathbf{W}_0\mathbf{\Gamma})(\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma})'^{-1}. \tag{9.12}$$

9.1.3 Optimal Choice of the Distance Matrix

When the number of moment conditions (q) is equal to the number of parameters to be estimated (p), the system is just identified. In the case of a just identified system, the GMM estimator does not depend on the choice of distance matrix. When $q > p$, there exist overidentifying restrictions and different GMM estimators are obtained for different distance matrices. In this case, one may choose the distance matrix that results in an (asymptotically) efficient GMM estimator. Hansen (1982) shows that the covariance matrix (9.12) is minimized when $\mathbf{W}_0 = \mathbf{\Omega}^{-1}$.[3] With this choice of the distance matrix, $\sqrt{T}(\mathbf{b}_T - \mathbf{b}_0)$ has an approximately normal distribution with mean zero and the covariance matrix

$$Cov(\mathbf{\Omega}^{-1}) = (\mathbf{\Gamma}'\mathbf{\Omega}^{-1}\mathbf{\Gamma})^{-1} \tag{9.13}$$

in large samples.

Let $\mathbf{\Omega}_T$ be a consistent estimator of $\mathbf{\Omega}$. Then $\mathbf{W}_T = \mathbf{\Omega}_T^{-1}$ is used to obtain $\mathbf{b}_T$. The resulting estimator is called the optimal or efficient GMM estimator. It should be noted, however, that it is optimal given $f(\mathbf{x}_t, \mathbf{b})$. In the context of instrumental variable estimation, this means that instrumental variables are given. The optimal selection of instrumental variables is discussed below in Section 9.7. Let $\mathbf{\Gamma}_T$ be a consistent estimator of $\mathbf{\Gamma}$. Then the standard errors of the optimal GMM estimator

[3] The covariance matrix is minimized in the sense that $Cov(\mathbf{W}_0) - Cov(\mathbf{\Omega}^{-1})$ is a positive semidefinite matrix for any positive definite matrix $\mathbf{W}_0$.

$\mathbf{b}_T$ are calculated as square roots of the diagonal elements of $\frac{1}{T}(\mathbf{\Gamma}'_T\mathbf{\Omega}_T^{-1}\mathbf{\Gamma}_T)^{-1}$. The appropriate method for estimating $\mathbf{\Omega}$ depends on the model. This problem is discussed in Chapter 6. It is usually easier to estimate $\mathbf{\Gamma}$ by $\mathbf{\Gamma}_T = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(\mathbf{x}_t,\mathbf{b}_T)}{\partial \mathbf{b}'}$ than to estimate $\mathbf{\Omega}$. In linear models, or in some simple nonlinear models, analytical derivatives are readily available. In nonlinear models, numerical derivatives are often used.

9.1.4 A Chi-Square Test for the Overidentifying Restrictions

In the case where there are overidentifying restrictions ($q > p$), a chi-square statistic can be used to test the overidentifying restrictions. One application of this test is to test the validity of the moment conditions implied by Euler equations for optimizing problems of economic agents. This application is discussed in Section 9.5. Hansen (1982) shows that T times the minimized value of the objective function, $TJ_T(\mathbf{b}_T)$, has an (asymptotic) chi-square distribution with $q-p$ degrees of freedom if $\mathbf{W}_0 = \mathbf{\Omega}^{-1}$ in large samples. This test is sometimes called Hansen's J test.[4]

Kyungho needs to check this!

If we reject the overidentifying restrictions based on Hansen's J test, it can be interpreted in two different ways. If a model implies the moment restrictions, for example, Euler equation approach, rejection of J test means that the model is rejected. However, if instrumental variables are chosen with common sense, rejection of J test means that instrumental variables are inappropriately chosen.

9.2 Special Cases

This section shows how linear regressions and nonlinear instrumental variable estimation are embedded in the GMM framework above.

[4]See Newey (1985) for an analysis of the asymptotic power properties of this chi-square test.

9.2.1 Ordinary Least Squares

Consider a linear model,

$$y_t = \mathbf{x}_{2t}'\mathbf{b}_0 + \epsilon_t, \tag{9.14}$$

where y_t and ϵ_t are stationary and ergodic random variables, $\mathbf{x}_{2t}$ is a p-dimensional stationary and ergodic random vector. OLS estimation can be embedded in the GMM framework by letting $\mathbf{x}_t = (y_t, \mathbf{x}_{2t}')'$, $f(\mathbf{x}_t, \mathbf{b}) = \mathbf{x}_{2t}(y_t - \mathbf{x}_{2t}'\mathbf{b})$, $\mathbf{u}_t = \mathbf{x}_{2t}\epsilon_t$, and $p = q$. Thus, the moment conditions (9.7) become the orthogonality conditions:

$$E(\mathbf{x}_{2t}\epsilon_t) = \mathbf{0}. \tag{9.15}$$

Since this is the case in a just identified system, the distance matrix $\mathbf{W}_0$ does not matter. Note that the OLS estimator minimizes $\sum_{t=1}^{T}(y_t - \mathbf{x}_{2t}'\mathbf{b})^2$ while the GMM estimator minimizes $(\sum_{t=1}^{T}\mathbf{x}_{2t}(y_t - \mathbf{x}_{2t}'\mathbf{b}))'(\sum_{t=1}^{T}\mathbf{x}_{2t}(y_t - \mathbf{x}_{2t}'\mathbf{b}))$. In this case, the GMM estimator coincides with the OLS estimator. To see this, note that $(\sum_{t=1}^{T}\mathbf{x}_{2t}(y_t - \mathbf{x}_{2t}'\mathbf{b}))'(\sum_{t=1}^{T}\mathbf{x}_{2t}(y_t - \mathbf{x}_{2t}'\mathbf{b}))$ can be minimized by setting $\mathbf{b}_T$ so that $\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}) = \mathbf{0}$ in the case of a just identified system. This result implies that $\sum_{t=1}^{T}\mathbf{x}_{2t}y_t = (\sum_{t=1}^{T}\mathbf{x}_{2t}\mathbf{x}_{2t}')\mathbf{b}_T$. Thus, as long as $\sum_{t=1}^{T}\mathbf{x}_{2t}\mathbf{x}_{2t}'$ is invertible, $\mathbf{b}_T = (\sum_{t=1}^{T}\mathbf{x}_{2t}\mathbf{x}_{2t}')^{-1}\sum_{t=1}^{T}\mathbf{x}_{2t}y_t$. Hence, the GMM estimator $\mathbf{b}_T$ coincides with the OLS estimator.

9.2.2 Linear Instrumental Variables Regressions

Consider the linear model (9.14) and let $\mathbf{z}_t$ be a q-dimensional random vector of instrumental variables. Then instrumental variable regressions are embedded in the GMM framework by letting $\mathbf{x}_t = (y_t, \mathbf{x}_{2t}', \mathbf{z}_t')'$, $f(\mathbf{x}_t, \mathbf{b}) = \mathbf{z}_t(y_t - \mathbf{x}_{2t}'\mathbf{b})$, and $\mathbf{u}_t = \mathbf{z}_t\epsilon_t$.

Thus, the moment conditions become the orthogonality conditions

$$E(\mathbf{z}_t \epsilon_t) = \mathbf{0}. \tag{9.16}$$

In the case of a just identified system ($q = p$), the instrumental variable regression estimator $(\sum_{t=1}^{T} \mathbf{z_t x}'_{2t})^{-1} \sum_{t=1}^{T} \mathbf{z_t} y_t$ coincides with the GMM estimator. For the case of an overidentified system ($q > p$), the two-stage least-squares estimators and the three-stage least-squares estimators (for multiple regressions) can be interpreted as optimal GMM estimators when ϵ_t is serially uncorrelated and conditionally homoskedastic.[5]

9.2.3 Linear GMM estimator

Consider the linear regression model (9.14). Let $\mathbf{z}_t$ be a q-dimensional random vector of instrumental variables, $\mathbf{x}_t = (y_t, \mathbf{x}'_{2t}, \mathbf{z}'_t)'$, $f(\mathbf{x}_t, \mathbf{b}) = \mathbf{z}_t(y_t - \mathbf{x}'_{2t}\mathbf{b})$, and $\mathbf{u}_t = \mathbf{z}_t \epsilon_t$.

For the case of an overidentified system ($q > p$), the linear GMM estimator, $\mathbf{b}_T$, is the solution of the minimization problem (9.9), where

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}) &= \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_t(y_t - \mathbf{x}'_{2t}\mathbf{b}) \\ &= \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_t y_t + \frac{1}{T}\sum_{t=1}^{T}(-\mathbf{z}_t \mathbf{x}'_{2t})\mathbf{b} \\ &\equiv \mathbf{s_{zy}} + \mathbf{\Gamma}_T \mathbf{b}, \end{aligned} \tag{9.17}$$

$\mathbf{s_{zy}}$ $(q \times 1)$ is the corresponding vector of sample moments of $E(\mathbf{z}_t y_t)$ and $\mathbf{\Gamma}_T$ $(q \times p)$ is the corresponding vector of sample moments of $E(\frac{\partial f(\mathbf{x}_t, \mathbf{b}_0)}{\partial \mathbf{b}'})$. The first order condition for the minimization problem with respect to $\mathbf{b}$ is

$$\mathbf{\Gamma}'_T \mathbf{W}_T \mathbf{\Gamma}_T \mathbf{b} = -\mathbf{\Gamma}'_T \mathbf{W}_T \mathbf{s_{zy}}, \tag{9.18}$$

[5] This interpretation can be seen by examining the first order condition for the minimization problem (9.9).

where $\mathbf{W}_T$ is a $(q \times q)$ positive semidefinite matrix satisfying equation (11.2). The linear GMM estimator, $\mathbf{b}_T$, can be obtained by multiplying both sides by the inverse of $\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T$:

$$\mathbf{b}_T(\mathbf{W}_T) = -(\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T)^{-1}\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{s}_{\mathbf{zy}}. \tag{9.19}$$

When there is a system of multiple linear equations, the multiple-equation GMM estimator can be obtained. Moreover, under the assumption of conditional homoskedasticity, the three-stage least-squares estimators can be shown to be a special case of multiple-equation GMM estimators. (For more detailed explanation, see Hayashi, 2000).

9.2.4 Nonlinear Instrumental Variables Estimation

GMM is often used in the context of nonlinear instrumental variable (NLIV) estimation. Chapter 10 presents some examples of applications based on the Euler equation approach. Let $g(\mathbf{x}_{1t}, \mathbf{b})$ be a k-dimensional vector of functions and $\epsilon_t = g(\mathbf{x}_{1t}, \mathbf{b}_0)$. Suppose that there exist conditional moment restrictions, $E[\epsilon_t|\mathrm{I}_t] = 0$. Here it is assumed that $\mathrm{I}_t \subset \mathrm{I}_{t+1}$ for any t. Let $\mathbf{z}_t$ be a $q \times k$ matrix of random variables that are in the information set I_t.[6] By the law of iterative expectations, we obtain the unconditional moment restrictions:

$$E[\mathbf{z}_t g(\mathbf{x}_{1t}, \mathbf{b}_0)] = \mathbf{0}. \tag{9.20}$$

Thus, we let $\mathbf{x}_t = (\mathbf{x}_{1t}', \mathbf{z}_t')'$ and $f(\mathbf{x}_t, \mathbf{b}) = \mathbf{z}_t g(\mathbf{x}_{1t}, \mathbf{b})$ in this case. Hansen (1982) points out that the NLIV estimators discussed by Amemiya (1974), Jorgenson and

[6] In some applications, $\mathbf{z}_t$ is a function of $\mathbf{b}$. This property does not cause any problems as long as the resulting $f(\mathbf{x}_t, \mathbf{b})$ can be written as a function of $\mathbf{b}$ and a stationary random vector $\mathbf{x}_t$.

Laffont (1974), and Gallant (1977) can be interpreted as optimal GMM estimators when ϵ_t is serially uncorrelated and conditionally homoskedastic.

Hansen and Singleton (1982) Consumption-Based Capital Asset Pricing Model (C-CAPM) can be an example of NLIV interpretation of GMM estimation. The Euler equation is

$$\frac{E[\beta c_{t+1}^{-\alpha} R_{t+1} | \mathrm{I}_t]}{c_t^{-\alpha}} = 1, \tag{9.21}$$

where R_{t+1} is the gross real return of any asset.[7] The observed c_t they use is obviously nonstationary, although the specific form of nonstationarity is not clear (difference stationary or trend stationary, for example). Hansen and Singleton use $\frac{c_{t+1}}{c_t}$ in their econometric formulation, which is assumed to be stationary.[8] Then we let $\mathbf{b}_0 = (\beta, \alpha)'$, $\mathbf{x}_{1t} = (\frac{c_{t+1}}{c_t}, R_{t+1})'$, and $g(\mathbf{x}_{1t}, \mathbf{b}_0) = \beta(\frac{c_{t+1}}{c_t})^{-\alpha} R_{t+1} - 1$.[9] Stationary variables in I_t, such as the lagged values of $\mathbf{x}_t$, are used for instrumental variables $\mathbf{z}_t$. In this case, $\mathbf{u}_t$ is in I_{t+1}, and hence $\mathbf{u}_t$ is serially uncorrelated.

9.3 Important Assumptions

This section discusses two assumptions under which large sample properties of GMM estimators are derived. These two assumptions are important in the sense that applied researchers have encountered cases where, unless special care is taken, these assumptions are obviously violated.

[7] This asset pricing equation can be applied to any asset returns. For example, Mark (1985) applies the Hansen-Singleton model in asset returns in foreign exchange markets.

[8] In the following, assumptions about trend properties of equilibrium consumption are made. The simplest model in which these assumptions are satisfied is a pure exchange economy, with the trend assumptions imposed on endowments.

[9] When multiple asset returns are used, $g(\mathbf{x}_t, \mathbf{b})$ becomes a vector of functions.

9.3.1 Stationarity

In Hansen (1982), $\mathbf{x}_t$ is assumed to be (strictly) stationary. Among other things, this assumption implies that when they exist, the unconditional moments $E(\mathbf{x}_t)$ and $E(\mathbf{x}_t\mathbf{x}'_{t+\tau})$ cannot depend on t for any τ. Thus, this assumption rules out deterministic trends, autoregressive unit roots, and unconditional heteroskedasticity. On the other hand, conditional moments $E(\mathbf{x}_{t+\tau}|\mathrm{I}_t)$ and $E(\mathbf{x}_{t+\tau}\mathbf{x}'_{t+\tau+s}|\mathrm{I}_t)$ can depend on I_t. Thus, the stationarity assumption does *not* rule out the possibility that $\mathbf{x}_t$ has conditional heteroskedasticity. It should be noted that it is not enough for $\mathbf{u}_t = f(\mathbf{x}_t, \mathbf{b}_0)$ to be stationary. It is required that $\mathbf{x}_t$ is stationary, so that $f(\mathbf{x}_t, \mathbf{b})$ is stationary for all admissible $\mathbf{b}$, not just for $\mathbf{b} = \mathbf{b}_0$ (see Section ??????????? for an example in which $f(\mathbf{x}_t, \mathbf{b}_0)$ is stationary but $f(\mathbf{x}_t, \mathbf{b})$ is not for other values of $\mathbf{b}$).

Masao needs to check this!

Gallant (1987) and Gallant and White (1988) show that the GMM strict stationarity assumption can be relaxed to allow for unconditional heteroskedasticity. This property does *not* mean that $\mathbf{x}_t$ can exhibit nonstationarity by having deterministic trends or autoregressive unit roots. Some of their regularity conditions are violated by these popular forms of nonstationarity. Recent papers by Andrews and McDermott (1995) and Dwyer (1995) show that the stationarity assumption can be further relaxed for some forms of nonstationarity. However, the long-run covariance matrix estimation procedure often needs to be modified to apply their asymptotic theory. For this reason, the strict stationarity assumption is emphasized in the context of time series applications rather than the fact that this assumption can be relaxed.

Since many macroeconomic variables exhibit nonstationarity, unless a researcher is careful this assumption can be easily violated in applications. As will be explained in Subsection 9.4.2, nonstationarity in the form of trend stationarity can be treated with ease. In order to treat another popular form of nonstationarity, unit-root nonstationarity, researchers have used transformations such as first differences or growth rates of variables (see Chapter 10 for examples).

9.3.2 Identification

Another important assumption of Hansen (1982) is related to identification. Let

$$J_0(\mathbf{b}) = \{E[f(\mathbf{x}_t, \mathbf{b})]\}'\mathbf{W}_0\{E[f(\mathbf{x}_t, \mathbf{b})]\}. \tag{9.22}$$

The identification assumption is that $\mathbf{b}_0$ is the unique minimizer of $J_0(\mathbf{b})$. Since $J_0(\mathbf{b}) \geq 0$ and $J_0(\mathbf{b}_0) = 0$, $\mathbf{b}_0$ is a minimizer. Hence, this assumption requires $J_0(\mathbf{b})$ to be strictly positive for any other $\mathbf{b}$. This assumption is obviously violated if $f(\mathbf{x}_t, \mathbf{b}) \equiv \mathbf{0}$ for some $\mathbf{b}$ that does not have any economic meaning (see Chapter 10 for examples). Even when this assumption is not violated, if values of $J_0(\mathbf{b})$ are close to zero for parameter values around the unique minimizer and for other parameter values, then we have *weak identification* problem. This problem will be discussed later in this chapter.

9.4 Extensions

This section explains econometric methods that are closely related to the basic GMM framework.

9.4.1 Sequential Estimation

This subsection discusses sequential estimation (or two step estimation). Consider a system

$$f(\mathbf{x}_t, \mathbf{b}) = \begin{bmatrix} f_1(\mathbf{x}_t, \mathbf{b}_1) \\ f_2(\mathbf{x}_t, \mathbf{b}_1, \mathbf{b}_2) \end{bmatrix}, \tag{9.23}$$

where $\mathbf{b} = (\mathbf{b}_1', \mathbf{b}_2')'$, $\mathbf{b}_i$ is a p_i-dimensional vector of parameters, and f_i is a q_i-dimensional vector of functions. Although it is possible to estimate $\mathbf{b}_1$ and $\mathbf{b}_2$ simultaneously, it may be computationally convenient to estimate $\mathbf{b}_1$ from $f_1(\mathbf{x}_t, \mathbf{b}_1)$ first, and then estimate $\mathbf{b}_2$ from $f_2(\mathbf{x}_t, \mathbf{b}_1, \mathbf{b}_2)$ in a second step (see, e.g., Barro, 1976; Atkeson and Ogaki, 1996, for examples of empirical applications). In general, the asymptotic distribution of the estimator of $\mathbf{b}_2$ is affected by the estimation of $\mathbf{b}_1$ (see, e.g., Newey, 1984; Pagan, 1984, 1986). A GMM computer program for sequential estimation can be used to calculate the correct standard errors that take into account these effects from estimating $\mathbf{b}_1$. If there are overidentifying restrictions in the system, an econometrician may wish to choose the second step distance matrix in an efficient way. The choice of the second step distance matrix is analyzed by Hansen, Heaton, and Ogaki (1992).

Suppose that the first step estimator $\mathbf{b}_T^1$ minimizes

$$J_{1T}(\mathbf{b}_1) = \{\frac{1}{T}\sum_{t=1}^{T} f_1(\mathbf{x}_t, \mathbf{b}_1)\}'\mathbf{W}_{1T}\{\frac{1}{T}\sum_{t=1}^{T} f_1(\mathbf{x}_t, \mathbf{b}_1)\} \tag{9.24}$$

and that the second step estimator minimizes

$$J_{2T}(\mathbf{b}_2) = \{\frac{1}{T}\sum_{t=1}^{T} f_2(\mathbf{x}_t, \mathbf{b}_{1T}, \mathbf{b}_2)\}'\mathbf{W}_{2T}\{\frac{1}{T}\sum_{t=1}^{T} f_2(\mathbf{x}_t, \mathbf{b}_{1T}, \mathbf{b}_2)\}, \tag{9.25}$$

where $\mathbf{W}_{iT}$ is a positive definite matrix that converges to $\mathbf{W}_{i0}$ with probability one. Let $\mathbf{\Gamma}_{ij}$ be the $q_i \times p_j$ matrix $E(\frac{\partial f_i}{\partial \mathbf{b}_j'})$ for $i = 1, 2$ and $j = 1, 2$.

Given an arbitrary $\mathbf{W}_{10}$, the optimal choice of the second step distance matrix is $\mathbf{W}_{20} = \mathbf{\Omega}^{*-1}$, where

$$\mathbf{\Omega}^* = [-\mathbf{\Gamma}_{21}(\mathbf{\Gamma}_{11}\mathbf{W}_{10}\mathbf{\Gamma}_{11})^{-1}\mathbf{\Gamma}_{11}\mathbf{W}_{10},\ \mathbf{I}]\ \mathbf{\Omega} \begin{bmatrix} -\mathbf{\Gamma}_{21}(\mathbf{\Gamma}_{11}\mathbf{W}_{10}\mathbf{\Gamma}_{11})^{-1}\mathbf{\Gamma}_{11}\mathbf{W}_{10} \\ \mathbf{I} \end{bmatrix}. \tag{9.26}$$

With this choice of $\mathbf{W}_{20}$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mathbf{b}_{2T} - \mathbf{b}_{20})$ has an (asymptotic) normal distribution with mean zero and the covariance matrix

$$(\boldsymbol{\Gamma}_{22}'\boldsymbol{\Omega}^{*-1}\boldsymbol{\Gamma}_{22})^{-1} \tag{9.27}$$

and $TJ_{2T}(\mathbf{b}_{2T})$ has an (asymptotic) chi-square distribution with $q_2 - p_2$ degrees of freedom. It should be noted that if $\boldsymbol{\Gamma}_{21} = \mathbf{0}$, then the effect of the first step estimation can be ignored because $\boldsymbol{\Omega}^* = \boldsymbol{\Omega}_{22} = E(f_2(\mathbf{x}_t, \mathbf{b}_0)f_2(\mathbf{x}_t, \mathbf{b}_0)')$.

9.4.2 GMM with Deterministic Trends

This subsection discusses how GMM can be applied to time series with deterministic trends (see Eichenbaum and Hansen, 1990; Ogaki, 1988, 1989, for empirical examples). Suppose that $\mathbf{x}_t$ is trend stationary rather than stationary. In particular, let

$$\mathbf{x}_t = d(t, \mathbf{b}_{10}) + \mathbf{x}_t^*, \tag{9.28}$$

where $d(t, \mathbf{b}_{10})$ is a function of deterministic trends such as time polynomials and $\mathbf{x}_t^*$ is detrended $\mathbf{x}_t$. Assume that $\mathbf{x}_t^*$ is stationary with $E(\mathbf{x}_t^*) = \mathbf{0}$ and that there are q_2 moment conditions

$$E(f_2(\mathbf{x}_t^*, \mathbf{b}_{10}, \mathbf{b}_{20})) = \mathbf{0}. \tag{9.29}$$

Let $\mathbf{b} = (\mathbf{b}_1', \mathbf{b}_2')'$, $f_1(\mathbf{x}_t, \mathbf{b}_1) = \mathbf{x}_t - d(t, \mathbf{b}_1)$ and $f(\mathbf{x}_t, \mathbf{b}) = [f_1(\mathbf{x}_t, \mathbf{b}_1)', f_2(\mathbf{x}_t^*, \mathbf{b}_1, \mathbf{b}_2)']'$. Then GMM can be applied to $f(\mathbf{x}_t, \mathbf{b})$ to estimate $\mathbf{b}_1$ and $\mathbf{b}_2$ simultaneously.

9.4.3 Other GMM Estimators

Several alternative estimators have been developed to deal with the poor small sample performance and weak identification problem of GMM.

One of them is the *continuous-updating estimator* provided by Hansen, Heaton, and Yaron (1996). It is obtained from changing the weighting matrix with each choice of the parameter instead of taking it as given in each step of GMM estimation. An advantage of this estimator is that it is invariant to how the moment conditions are scaled.

Others use the *information theoretic approach* to circumvent the need for estimating a weighting matrix in a two step GMM. They include the empirical likelihood estimator (see, e.g., Kitamura and Stutzer, 1997; Imbens, 1997, 2002; Imbens and Spady, 2002) and exponential tilting estimator (see, e.g., Imbens, Spady, and Johnson, 1998). These estimators are based on minimization of the Kullback-Leibler Information Criterion distance to estimate parameters and to test the over-identifying restrictions (see, e.g., Golan, 2002 for a recent explanation of information econometrics).

9.5 Hypothesis Testing and Specification Tests

This section discusses specification tests and Wald, Lagrange Multiplier (LM), and likelihood ratio type statistics for hypothesis testing. Gallant (1987), Newey and West (1987), and Gallant and White (1988) have considered these three test statistics, and Eichenbaum, Hansen, and Singleton (1988) considered the likelihood ratio type test for GMM (or a more general estimation method that includes GMM as a special case).

Consider s nonlinear restrictions

$$H_0 : R(\mathbf{b}_0) = \mathbf{r}, \tag{9.30}$$

where R is a $s \times 1$ vector of functions. The null hypothesis H_0 is tested against the alternative of $R(\mathbf{b}_0) \neq \mathbf{r}$. Let $\mathbf{\Lambda} = \frac{\partial R}{\partial \mathbf{b}'}|_{\mathbf{b}_0}$ and $\mathbf{\Lambda}_T$ be a consistent estimator for $\mathbf{\Lambda}$. It is assumed that $\mathbf{\Lambda}$ is of rank s. If the restrictions are linear, then $R(\mathbf{b}_0) = \mathbf{\Lambda}\mathbf{b}_0$ and

$\mathbf{\Lambda}$ is known. Let $\mathbf{b}_T^u$ be an unrestricted GMM estimator and $\mathbf{b}_T^r$ be a GMM estimator that is restricted by (9.30). It is assumed that $\mathbf{W}_0 = \mathbf{\Omega}^{-1}$ is used for both estimators.

The Wald test statistic is

$$(9.31) \qquad T(R(\mathbf{b}_T^u) - \mathbf{r})'[\mathbf{\Lambda}_T(\mathbf{\Gamma}_T'\mathbf{\Omega}_T^{-1}\mathbf{\Gamma}_T)^{-1}\mathbf{\Lambda}_T']^{-1}(R(\mathbf{b}_T^u) - \mathbf{r}),$$

where $\mathbf{\Omega}_T, \mathbf{\Gamma}_T$, and $\mathbf{\Lambda}_T$ are estimated from $\mathbf{b}_T^u$. The Lagrange multiplier test statistic is

$$(9.32) \quad LM_T = \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T^r)'\mathbf{\Omega}_T^{-1}\mathbf{\Gamma}_T\mathbf{\Lambda}_T'(\mathbf{\Lambda}_T\mathbf{\Lambda}_T')^{-1}[\mathbf{\Lambda}_T(\mathbf{\Gamma}_T'\mathbf{\Omega}_T^{-1}\mathbf{\Gamma}_T)^{-1}\mathbf{\Lambda}_T']^{-1}$$
$$(\mathbf{\Lambda}_T\mathbf{\Lambda}_T')^{-1}\mathbf{\Lambda}_T\mathbf{\Gamma}_T'\mathbf{\Omega}_T^{-1}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T^r),$$

where $\mathbf{\Omega}_T, \mathbf{\Gamma}_T$, and $\mathbf{\Lambda}_T$ are estimated from $\mathbf{b}_T^r$. Note that in linear models LM_T is equal to (9.31), where $\mathbf{\Omega}_T, \mathbf{\Gamma}_T$, and $\mathbf{\Lambda}_T$ are estimated from $\mathbf{b}_T^r$ rather than $\mathbf{b}_T^u$. (????? Need to reword) The likelihood ratio type test statistic is

Masao needs to check this!

$$(9.33) \qquad T(J_T(\mathbf{b}_T^r) - J_T(\mathbf{b}_T^u)),$$

which is T times the difference between the minimized value of the objective function when the parameters are restricted and the minimized value of the objective function when the parameters are unrestricted. It is important that the same estimator for $\mathbf{\Omega}$ is used for both unrestricted and restricted estimation for the likelihood ratio type test statistic. Under a set of regularity conditions, all three test statistics have asymptotic chi-square distributions with s degrees of freedom. The null hypothesis is rejected when these statistics are larger than the critical values obtained from chi-square distributions.

Existing Monte Carlo evidence suggests that the small sample distributions of the Lagrange multiplier test and the likelihood ratio type test are better approximated by their asymptotic distributions than those of the Wald test (see Gallant, 1987). Another disadvantage of the Wald test is that in general, the test result for nonlinear restrictions depends on the parameterization (see, e.g., Gregory and Veall, 1985; Phillips and Park, 1988).

Though the chi-square test for the overidentifying restrictions discussed in Section 9.1 has been frequently used as a specification test in applications of GMM, other specification tests applicable to GMM are available. These include tests developed by Singleton (1985), Andrews and Fair (1988), Hoffman and Pagan (1989), Andrews (1991), Ghysels and Hall (1990a,b,c), Hansen (1990), and Dufour, Ghysels, and Hall (1994). Some of these tests are discussed by Hall (1993).

9.6 Numerical Optimization

For nonlinear models, it is usually necessary to apply a numerical optimization method to compute a GMM estimator by numerically minimizing the criterion function, $J_T(\mathbf{b})$. The Newton-Raphson method (see, e.g., Hamilton, 1994, Chapter 5) is often used with an approximation method to calculate the Hessian matrix. A problem with the Newton-Raphson method and other practical numerical optimization methods is that global optimization is not guaranteed. The GMM estimator is defined as a global minimizer of a GMM criterion function, and the proof of its asymptotic properties depends on this assumption. Therefore, the use of a local optimization method can result in an estimator that is not necessarily consistent and asymptotically normal.

If the criterion function and parameter space are convex, then the criterion function has a unique local minimum, which is also the global minimum. In this case, a local optimization algorithm started at any parameter values should be able to reach an approximate global minimum.

For nonconvex problems, however, there can be many local minima. For such problems, an algorithm called multi-start is often used for GMM applications. In this algorithm, one starts a local optimization algorithm from initial values of the parameters to converge to a local minimum, and then one repeats the process a number of times with different initial values. The estimator is taken to be the parameter values that correspond to the smallest value of the criterion function obtained during the multi-start process.

It should be noted that this multi-start algorithm is used for a given distance matrix. When the two stage or iterative GMM estimators are used, a different distance matrix is used in each stage, and hence a different criterion function is minimized. In most GMM programs, one needs to save the distance matrix in a file in order to apply the multi-start algorithm in each stage.

A problem with the multi-start algorithm, however, is that it does not necessarily find the global optimum. Therefore, the estimator it delivers is not necessarily consistent and asymptotically normal. Andrews (1997) proposes a simple stopping-rule procedure that overcomes this difficulty.

9.7 The Optimal Choice of Instrumental Variables

In the NLIV model discussed in Section 9.2, there are infinitely many possible instrumental variables because any variable in I_t can be used as an instrument. Hansen (1985) characterizes an efficiency bound (that is, a greatest lower bound) for the asymptotic covariance matrices of the alternative GMM estimators and optimal instruments that attain the bound. Since it can be time consuming to obtain optimal instruments, an econometrician may wish to compute an estimate of the efficiency bound to assess efficiency losses from using ad hoc instruments. Hansen (1985) also provides a method for calculating this bound for models with conditionally homoskedastic disturbance terms with an invertible MA representation.[10] Hansen, Heaton, and Ogaki (1988) extend this method to models with conditionally heteroskedastic disturbances and models with an MA representation that is not invertible.[11] Hansen and Singleton (1996) calculate these bounds and optimal instruments for a continuous time financial economic model.

9.8 Small Sample Properties

In most cases, the exact small sample properties cannot be derived for GMM estimators. Monte Carlo simulations have been conducted to study them for various nonlinear and linear models. Tauchen (1986) shows that GMM estimators and test statistics have reasonable small sample properties for data produced by simulations for a C-CAPM. Ferson and Foerster (1994) find similar results for a model of expected returns of assets as long as GMM is iterated for estimation of $\mathbf{\Omega}$. Kocherlakota (1990) uses preference parameter values of $\beta = 1.139$ and $\alpha = 13.7$ (in Section 9.1) in his simulations for a C- CAPM that is similar to the Tauchen's (1986) model. While these parameter values do not violate any theoretical restrictions for existence of an equilibrium, they are much larger than the estimates of these preference parameters by Hansen and Singleton (1982) and others. Kocherlakota (1990) shows that GMM estimators for these parameters are biased downward and the chi-square test for the

[10]Hayashi and Sims' (1983) estimator is applicable to this example.

[11]Heaton and Ogaki (1991) provide an algorithm to calculate efficiency bounds for a continuous time financial economic model based on the Hansen, Heaton, and Ogaki's (1988) method.

overidentifying restrictions tends to reject the null too frequently compared with its asymptotic size. Mao (1990) reports that the chi-square test overrejects for more conventional values of these preference parameters in his Monte Carlo simulations.

Tauchen (1986) investigates small sample properties of Hansen's (1985) optimal instrumental variable GMM estimators. He finds that the optimal estimators do not perform well in small samples as compared to GMM estimators with ad hoc instruments. Tauchen (1986) and Kocherlakota (1990) recommend a small number of instruments rather than a large number of instruments when ad hoc instruments are used.

In some applications, scaling factors are another factor to affect finite sample GMM estimates. For example, Ni (1997) demonstrates that finite sample estimates are sensitive to scaling factors, and some seemingly reasonable scaling factors systematically lead to spurious estimates. However, Hansen, Heaton, and Yaron's (1996) continuous updating estimator is not affected by scaling factors.

Arellano and Bond (1991) report Monte Carlo results on GMM estimators for dynamic panel data models. They report that the GMM estimators have substantially smaller variances than commonly used Anderson and Hsiao's (1981) estimators in their Monte Carlo experiments. They also report that the small sample distributions of the serial-correlation tests they study are well approximated by their asymptotic distributions.

A very important small sample problem is weak identification, which we will discuss in the next section.

9.9 Weak Identification

In many applications, the identification condition holds but is almost violated in the sense that the values of the objective function evaluated at certain parameter values other than the true values are very close to the minimized value. In such applications, we have a *weak* problem. In the context of linear IV or NLIV estimation, this is called the *weak instrument variables problem.*

Nelson and Startz (1990) perform Monte Carlo simulations to investigate small sample properties of linear instrumental variables regressions. They show that instrumental variables estimators have poor sample properties when the instruments are weakly correlated with explanatory variables. In particular, they find that the chi-square test tends to reject the null too frequently compared with its asymptotic distribution, and that t-ratios tend to be too large when the instrument is poor. Their results for t-ratios may seem counterintuitive because one might expect that the consequence of having a poor instrument would be a large standard error and a low t-ratio. Staiger and Stock (1997) show that when the instruments are weakly correlated with the endogenous regressors, conventional asymptotic distribution theory fails even if the sample size is large. These results may be expected to carry over to NLIV estimation.

In the context of two stage least squares, Staiger and Stock (1997) suggest that first stage F-statistics, which tests the hypothesis that the instruments do not enter the first stage regression, should be reported at a minimum. Stock and Yogo (2005) advocates a pre-test rule to only use two stage least squares t statistics when the first stage F statistic exceeds ten. One strategy which continually changes the instruments until the F-statistics is significant is criticized by Hall, Rudebusch, and Wilcox (1996) as it tends to make matters worse in the Monte Carlo simulations.

9.10 Identification Robust Methods

When GMM has a weak identification problem, the conventional GMM asymptotics fails to provide reliable inferences. One solution is to use identification robust methods, which does not rely on the identification assumption. These methods can be applied without using a pre-test rule such as Stock and Yogo's (2005). If confidence intervals or regions of parameters generated by identification robust methods are large, that indicates the presence of the weak identification problem.

Let $\boldsymbol{\theta}$ denote a p-dimensional vector of parameters to be estimated. Consider the k dimensional vector of moment restrictions

$$E(f_t(\boldsymbol{\theta})) = 0 \tag{9.34}$$

for $t = 1, \cdots, T$ which is assumed to be uniquely satisfied at θ_0. The objective function for the CUE is:

$$Q(\theta) = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\theta) \right)' \hat{V}_{ff}(\theta)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\theta) \right) \tag{9.35}$$

where $\hat{V}_{ff}(\theta)$ is a consistent estimator of the $k \times k$ covariance matrix $V_{ff}(\theta)$ of the moment vector.

In addition to the moment vector $f_t(\theta)$, consider also its derivative with respect to θ:

$$q_t(\theta) = \text{vec}\left(\frac{\partial f_t(\theta)}{\partial \theta'} \right)$$

and $q_T = \frac{1}{T}\sum_{t=1}^{T} q_t(\theta)$.

We assume that in the large sample, $f_t(\theta)$ and $q_t(\theta)$ satisfy

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} f_t(\theta) - E(f_t(\theta)) \\ q_t(\theta) - E(q_t(\theta)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \phi_f(\theta) \\ \phi_\theta(\theta) \end{pmatrix}$$

where $(\ \phi_f(\theta)' \quad \phi_\theta(\theta)'\ )'$ is a $k(p+1)$ dimensional normally distributed random process with mean zero and positive semi-definite $k(p+1) \times k(p+1)$ dimensional covariance matrix

$$V(\theta) = \lim_{T\to\infty} \text{var} \left(\begin{array}{c} \frac{1}{\sqrt{T}} \sum_{t=1}^T f_t(\theta) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T q_t(\theta) \end{array} \right) = \left(\begin{array}{cc} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{array} \right),$$

with $V_{\theta f}(\theta) = V_{f\theta}(\theta)' = (V_{\theta f,1}(\theta)' \cdots V_{\theta f,p}(\theta)')'$, $V_{\theta\theta}(\theta) = V_{\theta\theta,ij}(\theta)$, $i, j = 1, \cdots, p$ and $V_{ff}(\theta)$, $V_{\theta f,i}(\theta)$, $V_{\theta\theta,ij}(\theta)$ are $k \times k$ dimensional matrices for $i, j = 1, \cdots, p$.

The derivative estimator $q_T(\theta)$ is correlated with the average moment vector $f_T(\theta)$ since $V_{\theta f}(\theta) \neq 0$. The weak instrument robust statistics therefore use an alternative estimator of the derivative of the unconditional expectation of the Jacobian that is asymptotically uncorrelated with $f_T(\theta)$:

$$\begin{aligned} \hat{D}_T(\theta_0) = [&q_{1,T}(\theta_0) - \hat{V}_{\theta f,1}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0) \cdots \\ &q_{p,T}(\theta_0) - \hat{V}_{\theta f,p}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0)], \end{aligned}$$

where $\hat{V}_{\theta f,i}(\theta)$ are $k_f \times k_f$ estimators of the covariance matrices $V_{\theta f,i}(\theta)$, $i = 1, \cdots, p$, $\hat{V}_{\theta f}(\theta) = (\hat{V}_{\theta f,1}(\theta)' \cdots \hat{V}_{\theta f,p}(\theta)')'$ and $q_T(\theta_0) = (1'_{1,T}(\theta_0) \cdots q'_{p,T}(\theta_0))'$.

The weak instrument robust statistics can be used for hypothesis testing on both subsets and the entire vector of the parameters. Let $\theta = (\alpha' : \beta')'$, with α and β being p_α and p_β dimensional vectors, respectively, such that $p_\alpha + p_\beta = p$. For tests on the entire set of parameters, consider $\beta = \theta$. Below, we introduce four statistics that test the hypothesis $H_0 : \beta = \beta_0$.

- The S-statistic of Stock and Wright (2000):

$$S(\beta_0) = Q(\tilde{\alpha}(\beta_0), \beta_0),$$

where $\tilde{\alpha}(\beta_0)$ is the CUE of α given that $\beta = \beta_0$. This is the CUE objective function (16.64).

- The score or Lagrange Multiplier statistic:

$$LM(\beta_0) = f_T(\tilde{\alpha}(\beta_0), \beta_0)'\hat{V}_{ff}(\tilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} P_{\hat{V}_{ff}(\tilde{\alpha}(\beta_0).\beta_0)^{-\frac{1}{2}}\hat{D}_T(\tilde{\alpha}(\beta_0),\beta_0)} \hat{V}_{ff}(\tilde{\alpha}(\beta_0), \beta_0)^{-\frac{1}{2}} f_T(\tilde{\alpha}(\beta_0), \beta$$

where $P_A \equiv A(A'A)^{-1}A'$ for a full rank matrix A. This can be considered as the inverse of the conditional information matrix (Kleibergen, 2007).

- The over-identification statistic:

$$SL(\beta_0) = S(\beta_0) - LM(\beta_0)$$

- The conditional likelihood ratio statistic:

$$CLR(\beta_0) = \frac{1}{2}\left[S(\beta_0) - rk(\beta_0) + \sqrt{\{S(\beta_0) + rk(\beta_0)\}^2 - 4SL(\beta_0)rk(\beta_0)}\right]$$

where $rk(\beta_0)$ is a statistic that tests for a lower rank value of $J(\tilde{\alpha}(\beta_0), \beta_0)$ and is a function of $\hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)$ and $\hat{V}_{\theta\theta.f}(\tilde{\alpha}(\beta_0), \beta_0) = \hat{V}_{\theta\theta}(\tilde{\alpha}(\beta_0), \beta_0) - \hat{V}_{\theta f}(\tilde{\alpha}(\beta_0), \beta_0)\hat{V}_{ff}(\tilde{\alpha}(\beta_0), \beta_0)^{-1}$

$$rk(\beta_0) = \min_{\phi \in R^{p-1}} T(\ 1 \quad \phi\ )\hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)' \left[W'\hat{V}_{\theta\theta.f}(\tilde{\alpha}(\beta_0), \beta_0)W\right]^{-1} \hat{D}_T(\tilde{\alpha}(\beta_0), \beta_0)(\ 1 \quad \phi\ )'$$

where $W = (\ I_{p\alpha} \quad \phi'\ )' \otimes I_k$. The CLR statistic is a GMM extension of the conditional likelihood ratio statistic of Moreira (2003) for the linear instrumental variables regression model with one included endogenous variable.

Confidence sets for the parameter(s) β are obtained by inverting each of the identification-robust statistics (Zivot, Startz, and Nelson, 1998). The $(1-\alpha)100\%$ confidence bounds coincide with the intersection of the 1-α value of the test statistic with the $(1-\alpha)$ line. A $(1-\alpha)100\%$ level confidence set thus constructed contains all the values of β_0 for which the corresponding test of the hypothesis $H_0 : \beta = \beta_0$ does not reject H_0 at the $\alpha\%$ level of significance. When testing for more than one parameter jointly, these are the $1 - \alpha$ contours of the graph of the function $1 - p(\boldsymbol{\theta})$ where $p(\boldsymbol{\theta})$ is the p-value of a test of a joint null hypothesis on a vector of parameters $\boldsymbol{\theta}$. Projection based confidence sets for an element of $\boldsymbol{\theta}$ can be obtained from these joint confidence sets by projecting the widest range of the contours to the corresponding axis.

Kleibergen and Mavroeidis (2009) use the four test statistics to conduct inference on the parameters of the New Keynesian Phillips Curve (NKPC). They find evidence that forward-looking dynamics in inflation are statistically significant and dominate backward-looking dynamics. However, the confidence intervals for the backward-looking dynamics are too wide to draw any conclusion on its significance. Moreover, even though the slope of the NKPC is estimated to be positive, it is not significantly different from zero in any of the tests. These results confirm those of several authors who have reported empirical evidence that the NKPC is relatively flat and that its GMM estimation suffers from the weak identification problem (Mavroeidis, 2005; Nason and Smith, 2008). Kleibergen and Mavroeidis (2009) also find that, overall, the LR statistic is at least as powerful as other tests in the Monte Carlo simulations, and that it also yields the smallest confidence sets in their empirical applications.

Appendix

9.A Asymptotic Theory for GMM

This Appendix reviews proofs for the asymptotic properties of GMM.

9.A.1 Asymptotic Properties of Extremum Estimators

Many estimators are formed by minimizing or maximizing objective functions. These estimators are called extremum estimators, or optimization estimators. A GMM estimator is a special case of an extremum estimator. In this section, we prove the consistency of extremum estimators. The next section applies the results to GMM. Given $(\mathrm{S}, \mathcal{F}, Pr)$, let $\mathbf{x}$ be a m-vector of random variables, $\mathbf{b}$ be a p-vector of parameters, and $J_T(\mathbf{x}, \mathbf{b})$ be a sequence of real valued functions. We will often denote $J_T(\mathbf{x}, \mathbf{b})$ by $J_T(\mathbf{b})$. For GMM, $\mathbf{x}$ will be taken as $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2', \cdots \mathbf{x}_T')'$, so that m is T times the dimension of $\mathbf{x}_t$. Thus, we allow m to be a function of T. The parameter $\mathbf{b}$ is a member of a set $\mathcal{B} \subset \mathcal{R}^p$, and $\mathcal{B}$ is called the parameter space.

An important condition for the consistency of extremum estimators relies on the concept of almost sure uniform convergence. Consider a sequence of functions $g_T : \mathcal{R}^r \times \mathcal{B} \longmapsto \mathcal{R}^q$, such that $g_T : (\cdot, \mathbf{b})$ is measurable for each $\mathbf{b}$ in $\mathcal{B}$ and $f(\mathbf{z}, \mathbf{b})$ is continuous on $\mathcal{B}$ for each $\mathbf{z}$ in $\mathcal{R}^r$. Then g_T converges to a nonstochastic function $g_0(\mathbf{b})$ *almost surely uniformly* in $\mathbf{b} \in \mathcal{B}$ if there exists $F \in \mathcal{F}$ with $Pr(F) = 1$, such that given any $\epsilon > 0$, for each s in F there exists an integer $T(s, \epsilon)$ such that for all $T > T(s, \epsilon)$, $sup_{\mathcal{B}} |g_T(\mathbf{x}(s), \mathbf{b}) - g_0(\mathbf{b})| < \epsilon$. Here $|\cdot|$ denotes the Euclidean norm. In this section, we will require that the sequence of real-valued functions $J_T(\mathbf{x}, \mathbf{b})$ converges to a nonstochastic function $J_0(\mathbf{b})$ almost surely uniformly in $\mathbf{b} \in \mathcal{B}$. In the next section, we will require a sequence of vector-valued functions converges almost surely uniformly.

Consider the following set of assumptions:

Assumption 9.A.1 The parameter space $\mathcal{B}$ is a compact set in $\mathcal{R}^p$.

Assumption 9.A.2 $J_T(\mathbf{x}, \mathbf{b})$ is continuous in $\mathbf{b} \in \mathcal{B}$ for all $\mathbf{x}$ and is a measurable function of $\mathbf{x}$ for all $\mathbf{b} \in \mathcal{B}$.

Assumption 9.A.3 $J_T(\mathbf{x}, \mathbf{b})$ converges to a nonstochastic function $J_0(\mathbf{b})$ almost surely uniformly in $\mathbf{b} \in \mathcal{B}$.

Assumption 9.A.4 $J_0(\mathbf{b})$ attains a unique global minimum at $\mathbf{b}_0$.

Since $\mathcal{B}$ is a subset in $\mathcal{R}^k$, Assumption 9.A.1 is equivalent to assuming that $\mathcal{B}$ is closed and bounded. Define an extreme estimator, $\mathbf{b}_T$, as a value that satisfies

$$J_T(\mathbf{b}_T) = \min_{\mathbf{b} \in \mathcal{B}} J_T(\mathbf{b}).$$

A complication is that the minimizer may not be unique, and it is not easy to prove that $\mathbf{b}_T$ can be chosen in such a way that $\mathbf{b}_T(\mathbf{x})$ is measurable. Different solutions to this problem are possible. Here, we have adopted a set of assumptions that are stronger than the assumptions in Theorem 4.1.1 of Amemiya (1985) for the weak consistency of extremum estimators. Amemiya (1985) states that if $\mathbf{b}_T$ is not unique, it is possible to choose a value in such a way that $\mathbf{b}_T(\mathbf{x})$ is a measurable function of $\mathbf{x}$. Assuming that $\mathbf{b}_T(\mathbf{x})$ is chosen this way, we can prove the strong consistency of extremum estimators.

Theorem 9.A.1 *(Strong consistency of extremum estimators)* If Assumptions 9.A.1 - 9.A.4 are satisfied, then $\mathbf{b}_T$ converges almost surely to $\mathbf{b}_0$.

Proof Given any $\epsilon > 0$, let $\eta(\epsilon)$ an open ball with the center $\mathbf{b}_0$ and the radius ϵ. If $\eta(\epsilon)^c \cap \mathcal{B}$ is empty for all ϵ, the result is trivial. Suppose that $\eta(\epsilon)^c \cap \mathcal{B}$ is nonempty. Since $\eta(\epsilon)^c \cap \mathcal{B}$ is compact and $J_0(\mathbf{b})$ is continuous under our assumptions, $\min_{\mathbf{b} \in \eta(\epsilon)^c \cap \mathcal{B}} J_0(\mathbf{b})$ exits. Denote

$$\delta(\epsilon) = \min_{\mathbf{b} \in \eta(\epsilon)^c \cap \mathcal{B}} J_0(\mathbf{b}) - J_0(\mathbf{b}_0).$$

Since $J_T(\mathbf{x}, \mathbf{b})$ converges almost surely uniformly to $J_0(\mathbf{b})$, there exists $F \in \mathcal{F}, Pr(F) = 1$ such that for each s in F and all $T > T(s, \delta(\epsilon))$, $|J_T(\mathbf{b}) - J_0(\mathbf{b})| < \frac{\delta(\epsilon)}{2}$. For $\mathbf{b} = \mathbf{b}_T$, we have

$|J_T(\mathbf{b}_T) - J_0(\mathbf{b}_T)| < \frac{\delta(\epsilon)}{2}$, and hence $J_0(\mathbf{b}_T) < J_T(\mathbf{b}_T) + \frac{\delta(\epsilon)}{2}$. For $\mathbf{b} = \mathbf{b}_0$, we have $|J_T(\mathbf{b}_0) - J_0(\mathbf{b}_0)| < \frac{\delta(\epsilon)}{2}$ or $J_T(\mathbf{b}_0) < J_0(\mathbf{b}_0) + \frac{\delta(\epsilon)}{2}$. Since $\mathbf{b}_T$ minimizes $J_T(\mathbf{b})$ on $\mathcal{B}$, $J_T(\mathbf{b}_T) < J_0(\mathbf{b}_0) + \frac{\delta(\epsilon)}{2}$. Therefore, $J_0(\mathbf{b}_T) < J_0(\mathbf{b}_0) + \delta(\epsilon)$ for each s in F and all $T > T(s, \delta(\epsilon))$. It follows that $\mathbf{b}_T \in \eta(\epsilon)$ for each s in F and all $T > T(s, \delta(\epsilon))$. Since ϵ is arbitrary and $Pr(F) = 1$, it follows that $\mathbf{b}_T$ converges to $\mathbf{b}_0$ almost surely. ∎

9.A.2 Consistency of GMM Estimators

In this section, we apply Theorem 9.A.1 to GMM estimators. We construct the objective function $J_T(\mathbf{x}, \mathbf{b})$ from a stationary ergodic stochastic process $\mathbf{x}_t$ and a function $f : \mathcal{R}^r \times \mathcal{B} \longmapsto \mathcal{R}^q$ where q is greater than or equal to p. We will often denote $f(\mathbf{x}_t, \mathbf{b})$ by $f_t(\mathbf{b})$ or $f(\mathbf{b})$. We retain Assumption 9.A.1, and impose conditions on $\mathbf{x}_t$ and f to ensure Assumptions 9.A.2 - 9.A.4 are satisfied.

Assumption 9.A.5 $\{\mathbf{x}_t : t \geq 1\}$ is an r-vector stationary and ergodic process.

Assumption 9.A.6 $f(\cdot, \mathbf{b})$ is measurable for each $\mathbf{b}$ in $\mathcal{B}$ and $f(\mathbf{z}, \mathbf{b})$ is continuous on $\mathcal{B}$ for each $\mathbf{z}$ in $\mathcal{R}^r$.

Assumption 9.A.7 $E(|f(\mathbf{x}_1, \mathbf{b})|)$ exists and is finite for all $\mathbf{b} \in \mathcal{B}$ and $E(f(\mathbf{x}_1, \mathbf{b}_0)) = \mathbf{0}$.

Since $\mathbf{x}_t$ is stationary and ergodic, $f(\mathbf{x}_t, \mathbf{b})$ is also stationary and ergodic for each $\mathbf{b}$. Therefore, Assumption 9.A.7 can be stated with any $\mathbf{x}_t$ instead of $\mathbf{x}_1$.

Consider the following set of assumptions:

Assumption 9.A.8 $\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b})$ converges almost surely uniformly to $E(f(\mathbf{b}))$ in $\mathcal{B}$.

Since $f(\mathbf{x}_t, \mathbf{b})$ is stationary and ergodic with finite first moments for each $\mathbf{b}$, $\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b})$ converges almost surely to $E(f(\mathbf{b}))$ for each $\mathbf{b}$ in $\mathcal{B}$. Assumption 9.A.8 assumes that this convergence is uniform. A sufficient condition for this assumption will be given in the next section.

Assumption 9.A.9 $E(f(\mathbf{b}))$ has a unique zero value at $\mathbf{b}_0$.

Assumption 9.A.10 The sequence of random positive semidefinite matrices $\{\mathbf{W}_T : T \geq 1\}$ converges almost surely to a nonstochastic positive definite matrix $\mathbf{W}_0$.

Let $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2', \cdots, \mathbf{x}_T')'$, $J_T(\mathbf{x}, \mathbf{b}) = \{\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b})\}'\mathbf{W}_T\{\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b})\}$, and $J_0(\mathbf{b}) = E(f(\mathbf{x}_1))'\mathbf{W}_0 E(f(\mathbf{x}_1))$. Define a GMM, $\mathbf{b}_T$, as a value that satisfies

$$J_T(\mathbf{b}_T) = \min_{\mathbf{b} \in \mathcal{B}} J_T(\mathbf{b}).$$

As in Section 9.A.1, it is understood that if $\mathbf{b}_T$ is not unique, we appropriately choose a value in such a way that $\mathbf{b}_T(\mathbf{x})$ is a measurable function of $\mathbf{x}$.

Theorem 9.A.2 *(Strong consistency of GMM estimators)* If Assumption 9.A.1, 9.A.5 - 9.A.10 are satisfied, $\mathbf{b}_T$ converges almost surely to $\mathbf{b}_0$. ∎

It is easy to verify that Assumptions 9.A.5 - 9.A.10 imply Assumptions 9.A.2 - 9.A.4. Therefore, Theorem 9.A.1 implies Theorem 9.A.2.

9.A.3 A Sufficient Condition for the Almost Sure Uniform Convergence

We directly assumed the uniform convergence in Assumption 9.A.8. It is very difficult to confirm that this assumption is satisfied in most econometric models. Hence, it is important to investigate sufficient conditions for Assumption 9.A.8 Hansen (1982) provides an important sufficient condition based on a concept called the first moment continuity of f. This section proves that the first moment continuity implies Assumption 9.A.8.

The following notation is used for our continuity restriction:

$$(9.A.1) \quad Mod_f(\delta, \mathbf{b}) = sup\{|f(\mathbf{b}) - f(\mathbf{b}^*)| : \mathbf{b}^* \in \mathcal{B} \text{ and } |\mathbf{b} - \mathbf{b}^*| < \delta\}.$$

where $|\cdot|$ denotes the Euclidean norm. Since $\mathcal{B}$ is separable, a dense sequence $\{\mathbf{b}_j : j \geq 1\}$ can be used in place of $\mathcal{B}$ in evaluating the *supremum*. In this case, $Mod_f(\delta, \mathbf{b})$ is a random variable for each positive value of δ and each $\mathbf{b}$ in $\mathcal{B}$. Also, $Mod_f(\delta, \mathbf{b}) \geq Mod_f(\delta^*, \mathbf{b})$ if δ is greater than δ^*. Since $f(\cdot, \mathbf{b})$ is continuous,

$$(9.A.2) \qquad \lim_{\delta \to 0} Mod_f(\delta, \mathbf{b}) = 0 \text{ for all } s \in \mathrm{S} \text{ and all } \mathbf{b} \in \mathcal{B}.$$

A function f is *first-moment continuous* if for each $\mathbf{b} \in \mathcal{B}$,

$$(9.A.3) \qquad \lim_{\delta \to 0} E[Mod_f(\delta, \mathbf{b})] = 0.$$

A necessary and sufficient condition for f to be first-moment continuous is that for each $\mathbf{b} \in \mathcal{B}$, there exists $\delta > 0$ such that

$$E[Mod_f(\delta, \mathbf{b})] < \infty.$$

It is trivial to see that this condition is necessary. This condition is sufficient because $Mod_f(\delta, p)$ is decreasing in δ: the Dominated Convergence Theorem and (9.A.2) imply the first-moment continuity of f.

Assumption 9.A.8′ f is first-moment continuous.

Proposition 9.A.1 Under Assumptions 9.A.1, 9.A.5 - 9.A.7, Assumption 9.A.8′ implies that Assumption 9.A.8 is satisfied. Therefore, Assumption 9.A.8 for Theorem 9.A.2 can be replaced by Assumption 9.A.8′.

The proof of this proposition given here is a modified version of the proof of a closely related theorem in Hansen, Heaton, and Ogaki (1992). The proof is long and technical but is presented here for the econometric theory-oriented readers. We prepare for the proof by proving three lemmas. To prove this proposition, we use (i) pointwise continuity of $E(f)$, (ii) a pointwise Law of Large Numbers for $\frac{1}{T}\sum f_t(\mathbf{b})$ for each $\mathbf{b}$ in $\mathcal{B}$, and (iii) a pointwise Law of Large Numbers for $\frac{1}{T}\sum_t Mod_f(\delta, \mathbf{b})$ for each $\mathbf{b}$ in $\mathcal{B}$ and positive δ. As will be established in Lemma 9.A.1, (i) yields an approximation of the form:

Approximation 9.A.1 There is positive-valued function $\delta^*(\mathbf{b}, j)$ satisfying

$$|E[f(\mathbf{b}^*)] - E[f(\mathbf{b})]| < \frac{1}{j} \tag{9.A.4}$$

for all $\mathbf{b}^* \in \mathcal{B}$ such that $|\mathbf{b} - \mathbf{b}^*| < \delta^*(\mathbf{b}, j)$. ∎

As will be demonstrated in Lemma 9.A.2, (ii) provides an approximation of the form:

Approximation 9.A.2 There is an integer-valued function $T^*(s, \mathbf{b}, j)$ and an indexed set $\mathbf{\Lambda}^*(\mathbf{b}) \in \mathcal{F}$ such that $Pr\{\mathbf{\Lambda}^*(\mathbf{b})\} = 1$ and

$$|\frac{1}{T}\sum_{t=1}^{T}[f_t(s, \mathbf{b})] - E[f(\mathbf{b})]| < \frac{1}{j} \tag{9.A.5}$$

for all $T \geq T^*(s, \mathbf{b}, j)$, and $s \in \mathbf{\Lambda}^*(\mathbf{b})$. ∎

As will be shown in Lemma 9.A.3, (iii) yields an approximation of the form:

Approximation 9.A.3 There exists an integer-valued function $T^+(s, \mathbf{b}, j)$, a positive function $\delta^+(\mathbf{b}, j)$, and an indexed set $\mathbf{\Lambda}^+(\mathbf{b}) \in \mathcal{F}$ such that $Pr\{\mathbf{\Lambda}^+(\mathbf{b})\} = 1$ and

$$|\frac{1}{T}[f(\mathbf{b}) - f(\mathbf{b}^*)]| < \frac{1}{j} \tag{9.A.6}$$

for all $\mathbf{b}^* \in \mathcal{B}$ such that $|\mathbf{b} - \mathbf{b}^*| < \delta^+(\mathbf{b}, j), T \geq T^+(s, \mathbf{b}, j)$, and $s \in \mathbf{\Lambda}^+(\mathbf{b})$. ∎

Although the statements of these approximations require some cumbersome notation, we use this notation to monitor when sets and numbers depend on the underlying parameter values and approximation criteria ($\mathbf{b}$ and j). We will prove this theorem by showing that the assumption of a compact parameter space can be used to obtain an approximation that is uniform over the parameter space.

We now consider formally these inequalities. Lemma 9.A.1 establishes the continuity of $E(f)$.

Lemma 9.A.1 If Assumptions 9.A.1, 9.A.6, 9.A.7, 9.A.8′ are satisfied, then so is inequality (9.A.4).

Proof Since f is first-moment continuous, there is a function $\delta^*(\mathbf{b}, j)$ such that

$$E[Mod_f[\delta^*(\mathbf{b}, j), \mathbf{b}]] < \frac{1}{j}. \tag{9.A.7}$$

Note, however, that

$$\begin{aligned} |Ef(\mathbf{b}^*) - Ef(\mathbf{b})| &\leq E|f(\mathbf{b}^*) - f(\mathbf{b})| \\ &\leq E\{Mod_f[\delta^*(\mathbf{b}, j), \mathbf{b}]\} \\ &< \frac{1}{j} \end{aligned} \tag{9.A.8}$$

for all $\mathbf{b}^* \in \mathcal{B}$ such that $|\mathbf{b} - \mathbf{b}^*| < \delta^*(\mathbf{b}, j)$. ∎

For each element $\mathbf{b}$ in $\mathcal{B}$, $f(\mathbf{b})$ is a random variable with a finite absolute first moment. Thus the Law of Large Numbers applies pointwise as stated in the following lemma.

Lemma 9.A.2 If Assumptions 9.A.1, 9.A.6, and 9.A.7 are satisfied, then so is inequality (9.A.5).

Proof Since $\mathbf{x}_t$ is stationary and ergodic, $\{\frac{1}{T}\sum_{t=1}^{T}[f(\mathbf{b})] : T \geq 1\}$ converges to $E[f(\mathbf{b})]$ on a set $\mathbf{\Lambda}^*(\mathbf{b}) \in \mathcal{F}$ satisfying $Pr\{\mathbf{\Lambda}^*(\mathbf{b})\} = 1$. ∎

The Law of Large Numbers also applies to time series averages of $Mod_f(\delta, \mathbf{b})$. Since the mean of $Mod_f(\delta, \mathbf{b})$ can be made arbitrarily small by choosing δ to be small, we can control the local variation of time series averages of the random function f.

Lemma 9.A.3 If Assumptions 9.A.1, 9.A.5, 9.A.6, and 9.A.8′ are satisfied, then so is inequality (9.A.5).

Proof Since f is first-moment continuous, $Mod_f(\frac{1}{n}, \mathbf{b})$ has a finite first moment for some positive integer n. Since $\mathbf{x}_t$ is stationary and ergodic, $\{\frac{1}{T}\sum_{t=1}^{T}[Mod_f(\frac{1}{j}, \mathbf{b})] : T \geq 1\}$ converges to $E[Mod_f(\frac{1}{j}, \mathbf{b})]$ on a set $\mathbf{\Lambda}^+(\mathbf{b}, j)$ satisfying $Pr\{\mathbf{\Lambda}^+(\mathbf{b}, j)\} = 1$ for $j \geq n$. Let

$$\mathbf{\Lambda}^+(\mathbf{b}) = \bigcap_{j \geq n} \mathbf{\Lambda}^+(\mathbf{b}, j).$$

Then $\mathbf{\Lambda}^+(\mathbf{b})$ is measurable and $Pr\{\mathbf{\Lambda}^+(\mathbf{b})\} = 1$.

For each j, choose $\frac{1}{\delta^+(\mathbf{b},j)}$ to equal some integer greater than or equal to n such that

$$E\{Mod_f[\delta^+(\mathbf{b}, j), \mathbf{b}]\} < \frac{1}{2j}. \tag{9.A.9}$$

Since $\{\frac{1}{T}\sum_{t=1}^{T}\{Mod_f[\delta^+(\mathbf{b}, j), \mathbf{b}]\} : T \geq 1\}$ converges almost surely to $E\{Mod_f[\delta^+(\mathbf{b}, j), \mathbf{b}]\}$ on $\mathbf{\Lambda}^+(\mathbf{b})$, there exists an integer-valued function $T^+(s, \mathbf{b}, j)$ such that

$$|\frac{1}{T}\sum_{t=1}^{T}\{Mod_f[\delta^+(\mathbf{b}, j), \mathbf{b}]\} - E\{Mod_f[\delta^+(\mathbf{b}, j), \mathbf{b}]\}| < \frac{1}{2j} \tag{9.A.10}$$

for $T \geq T^+(s, \mathbf{b}, j)$. Therefore, $\frac{1}{T}\sum_{t=1}^{T}\{Mod_f[\delta^+(\mathbf{b}, j), \mathbf{b}]\} < \frac{1}{j}$. Since $\frac{1}{T}|\sum_{t=1}^{T}[f_t(\mathbf{b}) - f_t(\mathbf{b}^*)]| \leq \frac{1}{T}\sum_{t=1}^{T}\{Mod_f[\delta^+(\mathbf{b}, j), \mathbf{b}]\}$,

$$\frac{1}{T}|\sum_{t=1}^{T}[f_t(\mathbf{b}) - f_t(\mathbf{b}^*)]| < \frac{1}{j} \tag{9.A.11}$$

for all $\mathbf{b}^* \in \mathcal{B}$ such that $|\mathbf{b} - \mathbf{b}^*| < \delta^+(\mathbf{b}, j)$, $T \geq T^+(s, \mathbf{b}, j)$, $s \in \mathbf{\Lambda}^+(\mathbf{b})$, and $j \geq 1$. ∎

We now combine the conclusions from Lemmas 9.A.1 - 9.A.3 to prove Proposition 9.A.1. The idea is to exploit that fact that $\mathcal{B}$ is compact to move from pointwise to uniform convergence. Notice that in inequalities (9.A.4) - (9.A.6), $\mathbf{\Lambda}^+, \mathbf{\Lambda}^*, T^+$ and T^* all depend on $\mathbf{b}$. In the following proof, we will use compactness to show how the dependence on the parameter value can be eliminated.

Proof of Proposition 9.A.1 In the proof of this proposition, we use notation given in (9.A.4) - (9.A.6). Let

$$O(\mathbf{b}, n) = \{\mathbf{b}^* \in \mathcal{B} : |\mathbf{b} - \mathbf{b}^*| < \min\{\delta^*(\mathbf{b}, n), \delta^+(\mathbf{b}, n)\}\}. \tag{9.A.12}$$

Then for each $n \geq 1$,

$$\mathcal{B} = \bigcup_{\mathbf{b} \in \mathcal{B}} O(\mathbf{b}, n). \tag{9.A.13}$$

Since $\mathcal{B}$ is compact

$$\mathcal{B} = \bigcup_{J \geq 1}^{N(n)} O(\mathbf{b}_j, n), \tag{9.A.14}$$

where $N(n)$ is integer-valued and $\{\mathbf{b}_j : j \geq 1\}$ is a sequence in $\mathcal{B}$. Let

$$\mathbf{\Lambda} \equiv \bigcap_{j \geq 1} [\mathbf{\Lambda}^*(\mathbf{b}_j) \cap \mathbf{\Lambda}^+(\mathbf{b}_j)]. \tag{9.A.15}$$

Then $\mathbf{\Lambda} \in \mathcal{B}$ and $Pr(\mathbf{\Lambda}) = 1$. Let

$$\begin{aligned} T(s, n) \equiv\ & \max\{T^*(s, \mathbf{b}_1, n), T^*(s, \mathbf{b}_2, n), \cdots, T^*[s, \mathbf{b}_{N(n)}, n], \\ & T^+(s, \mathbf{b}_1, n), T^+(s, \mathbf{b}_2, n), \cdots, T^+[s, \mathbf{b}_{N(n)}, n]\}. \end{aligned} \tag{9.A.16}$$

For $T \geq T(s, n)$, inequalities (9.A.4)-(9.A.6) imply that

$$\begin{aligned} & |\frac{1}{T}\sum_{t=1}^{T}[f_t(\mathbf{b})] - E[f(\mathbf{b})]| \\ \leq\ & \frac{1}{T}|\sum_{t=1}^{T}[f_t(\mathbf{b}_j)] - \sum_{t=1}^{T}[f(\mathbf{b}_j)]| + |\frac{1}{T}\sum_{t=1}^{T}[f_t(\mathbf{b}_j)] - E[f(\mathbf{b}_j)]| + |E[f(\mathbf{b}_j)] - E[f(\mathbf{b})]| \\ <\ & \frac{3}{n}, \end{aligned} \tag{9.A.17}$$

where $\mathbf{b}_j$ is chosen so that $\mathbf{b} \in O(\mathbf{b}_j, n)$ for some $1 \leq j \leq N(n)$. Therefore, $\frac{1}{T}\sum_{t=1}^{T} f_t$ converges almost surely uniformly to $E(f)$. ∎

9.A.4 Asymptotic Distributions of GMM Estimators

This section proves the asymptotic normality of GMM estimators and then discusses the optimal GMM estimators. It is possible to utilize the asymptotic normality results for general extremum estimators such as Amemiya's (1985) Theorem 4.1.3 here. However, unlike with the consistency results, it is more convenient to exploit the particular structure of the GMM objective function for this proof.

Consider the following set of assumptions:

Assumption 9.A.11 $\{\mathbf{b}_T : T \geq 1\}$ converges almost surely to $\mathbf{b}_0$.

Assumption 9.A.12 $\mathbf{b}_0 \in \mathcal{B}^o \subset \mathcal{B} \subset \mathcal{R}^p$.

Assumption 9.A.13 $f(\cdot, \mathbf{b})$ is continuously differentiable with respect to $\mathbf{b}$ on $\mathcal{B}^o$ and the derivative $Df(\cdot, \mathbf{b})$ has finite first moments and is first moment continuous on $\mathcal{B}^o$.

Assumption 9.A.14 $\{\mathbf{W}_T : T \geq 1\}$ converges almost surely to a nonsingular matrix $\mathbf{W}_0$ of real numbers.

Assumption 9.A.15 $\{\mathbf{x}_t : t \geq 1\}$ is stationary and ergodic.

Assumption 9.A.16 $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\mathbf{b}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Omega} = \sum_{j=-\infty}^{\infty} E(f_t(\mathbf{x}_t, \mathbf{b}_0) f_{t-j}(\mathbf{x}_t, \mathbf{b}_0)')$.

Assumption 9.A.17 $E(Df(\mathbf{x}_t, \mathbf{b}_0))$ has rank p.

We denote $E(Df(\mathbf{x}_1, \mathbf{b}_0))$ by $\mathbf{\Gamma}$ and $\frac{1}{T}\sum_{t=1}^{T} Df(\mathbf{x}_t, \mathbf{b}_T)$ by $\mathbf{\Gamma}_T$.

Theorem 9.A.3 *(Asymptotic normality of GMM estimators)* If Assumptions 9.A.11 - 9.A.17 are satisfied, then

$$\sqrt{T}(\mathbf{b}_T - \mathbf{b}_0) \xrightarrow{D} N(\mathbf{0}, (\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Omega}\mathbf{W}_0\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma})^{-1}).$$

Proof Assumptions 9.A.11 and 9.A.12 imply there exists $F \in \mathcal{F}$, $Pr(F) = 1$ such that for any s in F there exists an integer $T(s)$ such that $\mathbf{b}_T \in \mathcal{B}^o$ for all $T \geq T(s)$. Going forward, we assume that $\mathbf{b}_T \in \mathcal{B}^o$. The first order condition for the minimization of the objective function is

$$\mathbf{\Gamma}_T'\mathbf{W}_T\{\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T)\} = \mathbf{0}. \tag{9.A.18}$$

Given $\mathbf{x}_t$, applying the Mean Value Theorem to each row of $\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T)$, we obtain

$$\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T) = \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_0) + \mathbf{\Gamma}_T^*(\mathbf{b}_T - \mathbf{b}_0), \tag{9.A.19}$$

where $\mathbf{\Gamma}_T^*$ is formed by evaluating each row of $\frac{1}{T}\sum_{t=1}^{T} Df(\mathbf{x}_t,\mathbf{b})$ at an intermediate vector between $\mathbf{b}_T$ and $\mathbf{b}_0$. Assumptions 9.A.11 - 9.A.13, and 9.A.15 imply that $\mathbf{\Gamma}_T^*$ converges almost surely to $\mathbf{\Gamma}$. Combining (9.A.18) and (9.A.19), we obtain

$$\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T^*(\mathbf{b}_T-\mathbf{b}_0) = -\mathbf{\Gamma}_T'\mathbf{W}_T\{\frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t,\mathbf{b}_0)\}. \tag{9.A.20}$$

$\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T^*$ converges almost surely to $\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma}$, which is nonsingular. Hence, for sufficiently large T, $\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T^*$ is nonsingular with probability one. When $\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T^*$ is nonsingular

$$\sqrt{T}(\mathbf{b}_T-\mathbf{b}_0) = -(\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T^*)^{-1}\mathbf{\Gamma}_T'\mathbf{W}_T\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(\mathbf{x}_t,\mathbf{b}_0)\}. \tag{9.A.21}$$

Since $(\mathbf{\Gamma}_T'\mathbf{W}_T\mathbf{\Gamma}_T^*)^{-1}\mathbf{\Gamma}_T'\mathbf{W}_T$ converges almost surely to $(\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'\mathbf{W}_0$, Assumption 9.A.16 implies the conclusion. ∎

We use the following two propositions to prove that the GMM estimator with $\mathbf{W}_0 = \mathbf{\Omega}^{-1}$ is the optimal GMM estimator when $\mathbf{\Omega}$ is nonsingular.

Proposition 9.A.2 Let $\mathbf{A}$ be a $q \times p$ matrix of rank p, then $\mathbf{M} = \mathbf{I}_q - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ is idempotent with rank $q-p$. ∎

Proposition 9.A.3 Let $\mathbf{A}$ and $\mathbf{C}$ be symmetric nonsingular matrices of the same size. Then $\mathbf{A} \geq \mathbf{C} \geq \mathbf{0}$ implies $\mathbf{A}^{-1} \leq \mathbf{C}^{-1}$. ∎

Note that Proposition 9.A.2 implies that $\mathbf{M}$ is positive semidefinite.

Assumption 9.A.18 $\mathbf{\Omega}$ is nonsingular.

Let $Cov(\mathbf{W}_0) = (\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Omega}\mathbf{W}_0\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{W}_0\mathbf{\Gamma})^{-1}$. $Cov(\mathbf{W}_0)$ is the covariance matrix of the GMM estimator associated with $\mathbf{W}_0$. In particular, $Cov(\mathbf{\Omega}^{-1}) = (\mathbf{\Gamma}'\mathbf{\Omega}^{-1}\mathbf{\Gamma})^{-1}$.

Theorem 9.A.4 *(Optimal GMM Estimators)* Suppose that Assumptions 9.A.11 - 9.A.18 are satisfied. Then $Cov(\mathbf{\Omega}^{-1}) \leq Cov(\mathbf{W}_0)$ for any $p \times p$ positive definite matrix $\mathbf{W}_0$.

Proof Since $\boldsymbol{\Omega}^{-1}$ is positive definite, there exists a nonsingular $p\times p$ matrix $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Lambda}'\boldsymbol{\Lambda}$. Then $\boldsymbol{\Omega} = \boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda}')^{-1}$. Let $\mathbf{A}_1 = \boldsymbol{\Lambda}\boldsymbol{\Gamma}$ and $\mathbf{A}_2 = \boldsymbol{\Lambda}'^{-1}\mathbf{W}_0\boldsymbol{\Gamma}$. Since $\mathbf{I} - \mathbf{A}_2(\mathbf{A}_2'\mathbf{A}_2)^{-1}\mathbf{A}_2'$ is positive semidefinite by Proposition 9.A.2, we have

$$\mathbf{A}_1'\mathbf{A}_1 \geq \mathbf{A}_1'\mathbf{A}_2(\mathbf{A}_2'\mathbf{A}_2)^{-1}\mathbf{A}_2'\mathbf{A}_1. \tag{9.A.22}$$

From Proposition 9.A.3, we obtain

$$(\mathbf{A}_1'\mathbf{A}_1)^{-1} \leq (\mathbf{A}_1'\mathbf{A}_2)^{-1}\mathbf{A}_2'\mathbf{A}_2(\mathbf{A}_2'\mathbf{A}_1)^{-1}. \tag{9.A.23}$$

Since $Cov(\boldsymbol{\Omega}^{-1}) = (\boldsymbol{\Gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma})^{-1} = (\mathbf{A}_1'\mathbf{A}_1)^{-1}$ and $Cov(\mathbf{W}_0) = (\mathbf{A}_1'\mathbf{A}_2)^{-1}\mathbf{A}_2'\mathbf{A}_2(\mathbf{A}_2'\mathbf{A}_1)^{-1}$, the conclusion follows from this inequality. ∎

The next theorem gives the asymptotic distribution of Hansen's J test statistic for the overidentifying restrictions.

Theorem 9.A.5 *(Hansen's J test)* Suppose that Assumptions 9.A.11 - 9.A.18 are satisfied and that $\mathbf{W}_0 = \boldsymbol{\Omega}^{-1}$. Then TJ_T converges in distribution to a chi-square random variable with $q-p$ degrees of freedom.

Proof From (9.A.19), and Theorem 9.A.2,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T) \xrightarrow{D} N(\mathbf{0}, \mathbf{V}) \tag{9.A.24}$$

where $\mathbf{V} = [\mathbf{I}_q - \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}']\boldsymbol{\Omega}[\mathbf{I} - (\boldsymbol{\Gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}']$. As in the proof of Theorem 9.A.4, let $\boldsymbol{\Lambda}$ be a nonsingular $p \times p$ matrix such that $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Lambda}'\boldsymbol{\Lambda}$. Then $\boldsymbol{\Omega} = \boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda}')^{-1}$, and

$$\frac{1}{\sqrt{T}}\boldsymbol{\Lambda}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T) \xrightarrow{D} N(\mathbf{0}, \mathbf{M}) \tag{9.A.25}$$

where $\mathbf{M} = \boldsymbol{\Lambda}[\boldsymbol{\Omega} - \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}']\boldsymbol{\Lambda}' = \mathbf{I} - \boldsymbol{\Lambda}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Lambda}']$ is a symmetric, idempotent matrix. The trace of $\mathbf{M}$ is $q-p$ because $tr(\mathbf{M}) = tr(\mathbf{I}_q) - tr\{\boldsymbol{\Lambda}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma})^{-1}\ \boldsymbol{\Gamma}'\boldsymbol{\Lambda}'\} = tr(\mathbf{I}_q) - tr\{\boldsymbol{\Gamma}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma})^{-1}\} = tr(\mathbf{I}_q) - tr(\mathbf{I}_p) = q - p$. Therefore, there exists a matrix $\mathbf{F}$ such that $\mathbf{F}'\mathbf{F} = \mathbf{F}\mathbf{F}' = \mathbf{I}$, and

$$\mathbf{M} = \mathbf{F}\begin{bmatrix} \mathbf{I}_{q-p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{F}'. \tag{9.A.26}$$

Hence, if $\mathbf{y} \sim N(\mathbf{0}, \mathbf{M})$, then $\mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{F}\mathbf{F}'\mathbf{y} \sim \chi^2(q-p)$. Since $\mathbf{y}'\mathbf{y}$ is a continuous function mapping $\mathcal{R}^q$ into $\mathcal{R}$,

$$\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T)'\}\boldsymbol{\Omega}_T^{-1}\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(\mathbf{x}_t, \mathbf{b}_T)\} \xrightarrow{D} \chi^2(q-p) \tag{9.A.27}$$

where $\boldsymbol{\Omega}_T$ is a weakly consistent estimator for $\boldsymbol{\Omega}$. ∎

9.B The Conditional Likelihood Ratio Statistic

The conditional likelihood ratio (CLR) statistic can be used for an identification robust method to solve weak identification problems as explained in the text. The CLR statistic was proposed by Moreira (2003) for the linear IV regression models and later extended to GMM by Kleibergen (2005).

Kleibergen (2005) proposes a GMM Lagrange multiplier statistic (the K statistic) whose asymptotic χ^2 distribution holds in a wider set of circumstances such as the presence of weak identification problem. The K statistic replaces the sample average of the derivatives of the moments in the Newey and West's (1987) GMM LM statistic with a Jacobian estimator based on the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996). The CUE, $\hat{\theta}$, is obtained by minimizing the objective function, $Q(\theta)$, and continuously altering the covariance matrix as $\hat{\theta}$ is changed in the minimization. Because of the correlation between the Jacobian estimator and the average moment vector, the limiting behavior of the Newey-West GMM LM statistic depends on nuisance parameters when, for example, the expected Jacobian is zero. The Jacobian estimator based on the CUE in the K statistic avoids this problem for it is asymptotically uncorrelated with the average moment vector (Brown and Newey, 1998; Donald and Newey, 2000). Give the dataset $Y = [Y_1 \dots Y_T]'$, the K statistic for testing $H_0 : \theta = \theta_0$ is

$$K(\theta_0) = \frac{1}{4T} \left(\frac{\partial Q(\theta)}{\partial \theta'} \bigg|_{\theta_0} \right) \left[\hat{D}_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} \hat{D}_T(\theta_0, Y) \right]^{-1} \left(\frac{\partial Q(\theta)}{\partial \theta'} \bigg|_{\theta_0} \right)'$$

where $\hat{D}_T(\theta_0, Y)$ is the CUE Jacobian estimator, and $\hat{V}_{ff}$ the positive definite covariance matrix of the vector function $f_T(\theta, Y)$, and has a $\chi^2(m)$ limiting distribution under H_0 and necessary assumptions.

By construction, the K statistic is equal to zero around the values of θ for which the objective function attains its minimum, maximum, or is at an inflection point. While the moment conditions are satisfied for the values of θ where the objective function is minimal and the CUE is obtained, they are not satisfied at the maximal value and inflection points, and thus the K statistic suffers from a spurious decline in power for such values of θ. In order to appropriately account for this spurious behavior of the K statistic, Kleibergen suggests applying a GMM extension of Moreira's (2003) conditional likelihood ratio statistic for linear instrumental variables regressions (Kleibergen, 2004). The K statistic is combined with a J statistic,

$$J(\theta_0) = \frac{1}{T} f_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1/2} M_{\hat{V}_{ff}(\theta_0)^{-1/2}\hat{D}_T(\theta_0,Y)} \hat{V}_{ff}(\theta_0)^{-1/2} f_T(\theta_0, Y)$$

which tests the validity of the moment equations and is asymptotically independent of the K statistic.[12] For these values of θ where the objective function is at its maxima or a reflection point, the J statistic has discriminatory power because it tests the validity of the moment equations, $H_m : E(f_t(\theta_0)) = 0$, while the K statistic tests $H_0 : \theta = \theta_0$ given that the moment equations hold (Kleibergen, 2004).

The resulting test statistic (the GMM-M statistic) which accounts for the spurious power decline is

$$\text{GMM-M}(\theta_0) = \frac{1}{2}\left\{K(\theta_0) + J(\theta_0) - \text{rk}(\theta_0) + \sqrt{[K(\theta_0) + J(\theta_0) + \text{rk}(\theta_0)]^2 - 4J(\theta_0)\text{rk}(\theta_0)}\right\}$$

where $\text{rk}(\theta_0)$ is a statistic for testing the hypothesis of a lower rank value of $J_\theta(\theta_0)$, $H_r : \text{rank}(J_\theta(\theta_0)) = m - 1$ as in Cragg and Donald (1996), Cragg and Donald (1997), Kleibergen and Paap (2006), and Robin and Smith (2000). The GMM-M(θ_0) leads

[12] $M_A = I_T - P_A$ where $P_A = A(A'A)^{-1}A'$ for a full rank matrix A.

to inference that is centered around $\hat{\theta}$ when GMM-M$(\hat{\theta}) = 0$. This occurs when rk$(\hat{\theta})$ exceeds $J(\hat{\theta})$ which puts a condition on the rank statistic rk(θ_0) to be used in the GMM-M statistic.

A confidence set for θ can be obtained by specifying sequences of n increasing values for every element of θ and creating an m-dimensional grid that contains n^m different values of θ_0. The statistic of interest (i.e., the J, K, or GMM-M statistic) can then be computed for each of these n^m different values of θ_0. All elements in the specified grid for which the asymptotic p-value of the statistic of interest exceeds α are in the $(1 - \alpha)100\%$ asymptotic confidence set.

9.C A Procedure for Hansen's J Test (GMM.EXP)

Hansen's J test proceeds as follows:

(i) Check whether the number of moment restrictions is greater than that of the estimated parameters (the corresponding condition in the program is `NMR > KGM`).

(ii) Choose an appropriate method to estimate the long-run covariance matrix, Ω_T. See chapter 6 for details (the corresponding variable to specify the method is `CALWFLAG`).

(iii) Set the maximum number of iterations to estimate the optimal weighting matrix, $W_T = \Omega_T^{-1}$ (the default is `MAXITEGM =5`).

(iv) Define the objective function (the corresponding part in the program to define the GMM disturbance is the `HU` procedure.)

(v) If the test statistic value (`CHI` in the output) is greater than the critical value for the significance level you have in mind, say 5%, then reject the null hypothesis that the over-identification restrictions are satisfied.

Exercises

The following problems are on econometric theory and require materials in Appendix 9.A.

9.1 *(The Minimum Distance Estimation)* Assume that the following set of assumptions is satisfied.

(A1) $\mathbf{p}_T$ converges almost surely to a k-dimensional vector $\mathbf{p}_0$ of real numbers.

(A2) $\sqrt{T}(\mathbf{p}_T - \mathbf{p}_0)$ converges in distribution to a normally distributed random vector with mean zero and a nonsingular covariance matrix $\mathbf{\Sigma}$.

(A3) $\mathbf{\Sigma}_T$ converges almost surely to $\mathbf{\Sigma}$.

(A4) $\mathbf{p}_0 = \phi(\mathbf{q}_0)$ where ϕ is a continuously differentiable function that maps $\mathrm{Q} \subset \mathrm{R}^h$ into R^k. The parameter space Q is assumed to be compact. Let $D\phi(\mathbf{q})$ be the $k \times h$ matrix of the derivative of ϕ, then $\mathbf{D}_0 = D\phi(\mathbf{q}_0)$ is assumed to be of rank h.

(A5) $\mathbf{p}_0 \neq \phi(\mathbf{q})$ for all $\mathbf{q}$ in Q except for $\mathbf{q} = \mathbf{q}_0$.

Consider estimating $\mathbf{q}_0$ by minimizing

$$J_T(\mathbf{q}) = \{[\phi(\mathbf{q}) - \mathbf{p}_T]\}'\mathbf{W}_T\{[\phi(\mathbf{q}) - \mathbf{p}_T]\} \tag{9.E.1}$$

over Q, where $\mathbf{W}_T$ is a positive semidefinite $k \times k$ random matrix that converges almost surely to a positive definite matrix of real numbers $\mathbf{W}$. Let $\mathbf{q}_T$ be the minimizer. The estimator $\mathbf{q}_T$ is called the minimum distance estimator. Suppose that the sequence of minimizers converges almost surely to $\mathbf{q}_0$.

(**a**) Prove that $\mathbf{q}_T$ is strongly consistent for $\mathbf{q}_0$ by applying Theorem 9.A.1 attached at the end. Hint: (i) You do not need the first moment continuity to prove the almost sure uniform convergence. (ii) Define a norm for a matrix $\mathbf{w}$ by $|\mathbf{w}| = |\text{vec}(\mathbf{w})|$. Then $|\mathbf{wz}| \leq |\mathbf{w}||\mathbf{z}|$ for two conformable matrices $\mathbf{w}$ and $\mathbf{z}$.

(**b**) Derive the asymptotic distribution of the estimators as a function of $\mathbf{W}$.

(**c**) Derive the greatest lower bound for the asymptotic covariance matrices of members of this family of estimators, using Propositions 9.A.2 and 9.A.3 attached at the end. What is the optimal $\mathbf{W}$?

(**d**) Let $\mathbf{q}_T$ be the minimum distance estimator associated with the optimum distance matrix $\mathbf{W}$ in $\mathcal{B}$. Show that the minimized value of $TJ_T(\mathbf{q}_T)$ converges in distribution to a χ^2 random variable. What is the degree of freedom of this χ^2 test statistic?

(**e**) Consider the model

(9.E.2) $$y_t = \mathbf{x}_t'\mathbf{p}_0 + \epsilon_t,$$

where y_t, $\mathbf{x}_t$ are a stationary and ergodic random variable and a 2-dimensional random vector with finite second moments, respectively. Suppose that $E(\mathbf{x}_t\epsilon_t) = \mathbf{0}$, and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\epsilon_t$ converges in distribution to $N(\mathbf{0}, \mathbf{\Omega})$. Suppose that economic theory imposes the restriction $p_{02} = (p_{01})^3$, where p_{0i} is the i-th element of $\mathbf{p}_0$. Discuss how you estimate this model, imposing the restriction using the minimum distance procedure you studied in the earlier parts of this problem, assuming that you get the initial estimator, $\mathbf{p}_T$, by unconstrained OLS. In particular, discuss how do you obtain an estimator for $\mathbf{\Sigma}$, $\mathbf{\Sigma}_T$, and how you attain the bound you derived.

(**f**) Derive the asymptotic variance of your efficient minimum distance estimator you studied in (**e**) in terms of $\mathbf{\Omega}$ and p_{01}.

9.2 In the case of a just identified system ($q = p$), show that the instrumental variable regression estimator $(\sum_{t=1}^{T} \mathbf{z_t}\mathbf{x}_{2t}')^{-1} \sum_{t=1}^{T} \mathbf{z_t} y_t$ coincides with the GMM estimator.

9.3 All files needed for this problem are in the GMM-CCR package. You need to use GMM and KPRGMM. Modify INDIVIS.G program (you will need to make minor modifications to the `bgm, nw`, the `nf, fc, fx, fe` in PROC INDIVIS, `mm` in PROC MOMETNTS, `dm` in PROC DATAMOM, and PROC HU procedures) as follows:
Use $f_t = i_t$ only.
Estimate only 9 parameters ($\theta, A_a, \rho_a, \sigma_a, A_y, \log\gamma, \delta, \alpha$, and σ_i).

(**a**) Compute GMM estimates and standard errors of the above nine parameters.

(**b**) Compute the model moment of investment (σ_i) with its standard errors, and the data moment of investment (σ_i) with its standard errors

(**c**) Report the Wald test statistics and p-value to compare these two numbers.

References

Amemiya, T. (1974): "The Nonlinear Two-Stage Least -Squares Estimator," *Journal of Econometrics*, 2, 105–110.

——— (1985): *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts.

Anderson, T. W., and C. Hsiao (1981): "Estimation of Dynamic-Models with Error-Components," *Journal of the American Statistical Association*, 76(375), 598–606.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

——— (1997): "A Stopping Rule for the Computation of Generalized Method of Moments Estimators," *Econometrica*, 65, 913–931.

Andrews, D. W. K., and R. C. Fair (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55(4), 615–639.

Andrews, D. W. K., and C. J. McDermott (1995): "Nonlinear Econometric Models with Deterministically Trending Variables," *Review of Economic Studies*, 62, 343–360.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58(2), 277–297.

Atkeson, A., and M. Ogaki (1996): "Wealth-Varying Intertemporal Elasticities of Substitution: Evidence from Panel and Aggregate Data," *Journal of Monetary Economics*, 38, 507–534.

Barro, R. J. (1976): "Rational Expectations and the Role of Monetary Policy," *Journal of Monetary Economics*, 2, 1–32.

Brown, B. W., and W. K. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66(2), 453–464.

Cragg, J. G., and S. G. Donald (1996): "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association*, 91, 1301–1309.

——— (1997): "Inferring the Rank of a Matrix," Journal of Econometrics," *Journal of Econometrics*, 76(1–2), 223–250.

Donald, S. G., and W. K. Newey (2000): "A Jackknife Interpretation of the Continuous Updating Estimator," *Economics Letters*, 67(3), 239–243.

Dufour, J.-M., E. Ghysels, and A. Hall (1994): "Generalized Predictive Tests and Structural Change Analysis in Econometrics," *International Economic Review*, 35(1), 199–229.

Dwyer, M. (1995): "Essays in Nonlinear, Nonstationary, Time Series Econometrics," Ph.D. thesis, University of Rochester.

Eichenbaum, M., and L. P. Hansen (1990): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business and Economic Statistics*, 8, 53–69.

Eichenbaum, M., L. P. Hansen, and K. J. Singleton (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103, 51–78.

Ferson, W. E., and S. R. Foerster (1994): "Finite Sample Properties of the Generalized Methods of Moments in Tests of Conditional Asset Pricing Models," *Journal of Financial Economics*, 36, 29–55.

Gallant, A. R. (1977): "Three-stage Least-squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," *Journal of Econometrics*, 5, 71–88.

——— (1987): *Nonlinear Statistical Models*. John Wiley and Sons, New York.

Gallant, A. R., and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, New York.

Ghysels, E., and A. Hall (1990a): "Are Consumption-Based Intertemporal Capital Asset Pricing Models Structural?," *Journal of Econometrics*, 45, 121–139.

——— (1990b): "A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator," *International Economic Review*, 31, 355–364.

——— (1990c): "Testing Nonnested Euler Conditions with Quadratic-Based Methods of Approximation," *Journal of Econometrics*, 46, 273–308.

Golan, A. (2002): "Information and Entropy Econometrics–Editor's View," *Journal of Econometrics*, 107(1–2), 1–15.

Gregory, A. W., and M. R. Veall (1985): "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 53(6), 1465–1468.

Hall, A. (1993): "Some Aspects of Generalized Method of Moments Estimation," in *Handbook of Statistics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, vol. 11, chap. 15, pp. 393–417. Elsevier Science Publishers.

Hall, A. R., G. D. Rudebusch, and D. W. Wilcox (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37(2), 283–298.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.

Hansen, B. E. (1990): "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models," Manuscript, University of Rochester.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

——— (1985): "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30, 203–238.

Hansen, L. P., J. C. Heaton, and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, 83(403), 863–871.

——— (1992): "Lecture notes on GMM," Lecture notes.

Hansen, L. P., J. C. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), 262–280.

Hansen, L. P., and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(5), 1269–1286.

——— (1996): "Efficient Estimation of Linear Asset-Pricing Models with Moving Average Errors," *Journal of Business and Economic Statistics*, 14, 53–68.

Hayashi, F. (2000): *Econometrics*. Princeton University Press, Princeton.

Hayashi, F., and C. A. Sims (1983): "Nearly Efficient Estimation of Time Series Models with Predetermined, but Not Exogenous Instruments," *Econometrica*, 51, 783–798.

Heaton, J. C., and M. Ogaki (1991): "Efficiency Bound Calculations for a Time Series Model with Conditional Heteroskedasticity," *Economics Letters*, 35, 167–171.

Hoffman, D. L., and A. R. Pagan (1989): "Post-Sample Prediction Tests for Generalized-Method of Moments Estimators," *Oxford Bulletin of Economics and Statistics*, 51(3), 333–343.

Imbens, G. W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economic Studies*, 64(3), 359–383.

——— (2002): "Generalized Method of Moments and Empirical Likelihood," *Journal of Business and Economic Statistics*, 20(4), 493–506.

Imbens, G. W., and R. H. Spady (2002): "Confidence Intervals in Generalized Method of Moments Models," *Journal of Econometrics*, 107(1–2), 87–98.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66(2), 333–358.

Jorgenson, D. W., and J.-J. Laffont (1974): "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances," *Annals of Economic and Social Measurement*, 3, 615–640.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65(4), 861–874.

Kleibergen, F. (2004): "Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model," *Review of Economics and Statistics*, 86(1), 418–423.

——— (2005): "Testing Parameters in GMM without Assuming That They Are Identified," *Econometrica*, 73(4), 1103–1124.

——— (2007): "Generalizing Weak Instrument Robust IV Statistics towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, 139(1), 181–216.

Kleibergen, F., and S. Mavroeidis (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics*, 27(3), 293–311.

Kleibergen, F., and R. Paap (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133(1), 97–126.

Kocherlakota, N. R. (1990): "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.

Mao, C. S. (1990): "Hypothesis Testing and Finite Sample Properties of Generalized Method of Moments Estimators: A Monte Carlo Study," Working Paper No. 90–12, Federal Reserve Bank of Richmond.

Mark, N. C. (1985): "On Time Varying Risk Premia in the Foreign Exchange Market: An Econometric Analysis," *Journal of Monetary Economics*, 16, 3–18.

MAVROEIDIS, S. (2005): "Identification Issues in Forward-Looking Models Estimated by GMM with an Application to the Phillips Curve," *Journal of Money, Credit, and Banking*, 37(3), 421–448.

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

NASON, J. M., AND G. W. SMITH (2008): "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics*, 23(5), 525–551.

NELSON, C. R., AND R. STARTZ (1990): "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One," *Journal of Business*, 63, S125–S140.

NEWEY, W. K. (1984): "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters*, 14(2–3), 201–206.

——— (1985): "Generalized-Method of Moments Specification Testing," *Journal of Econometrics*, 29(3), 229–256.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

NI, S. (1997): "Scaling Factors in Estimation of Time-Nonseparable Utility Functions," *Review of Economics and Statistics*, 79(2), 234–240.

OGAKI, M. (1988): "Learning about Preferences from Time Trends," Ph.D. thesis, University of Chicago.

——— (1989): "Information in Deterministic Trends about Preferences," Manuscript.

PAGAN, A. R. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247.

——— (1986): "Two Stage and Related Estimators and Their Applications," *Review of Economic Studies*, 53, 517–538.

PHILLIPS, P. C. B., AND J. Y. PARK (1988): "Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors," *Journal of the American Statistical Association*, 83, 111–115.

ROBIN, J.-M., AND R. J. SMITH (2000): "Tests of Rank," *Econometric Theory*, 16(2), 151–175.

SINGLETON, K. J. (1985): "Testing Specifications of Economic Agents' Intertemporal Optimum Problems in the Presence of Alternative Models," *Journal of Econometrics*, 30(1–2), 391–413.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

STOCK, J. H., AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock, pp. 80–108. Cambridge University Press, Cambridge.

Tauchen, G. (1986): "Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics*, 4, 397–416.

Zivot, E., R. Startz, and C. R. Nelson (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39(4), 1119–1144.