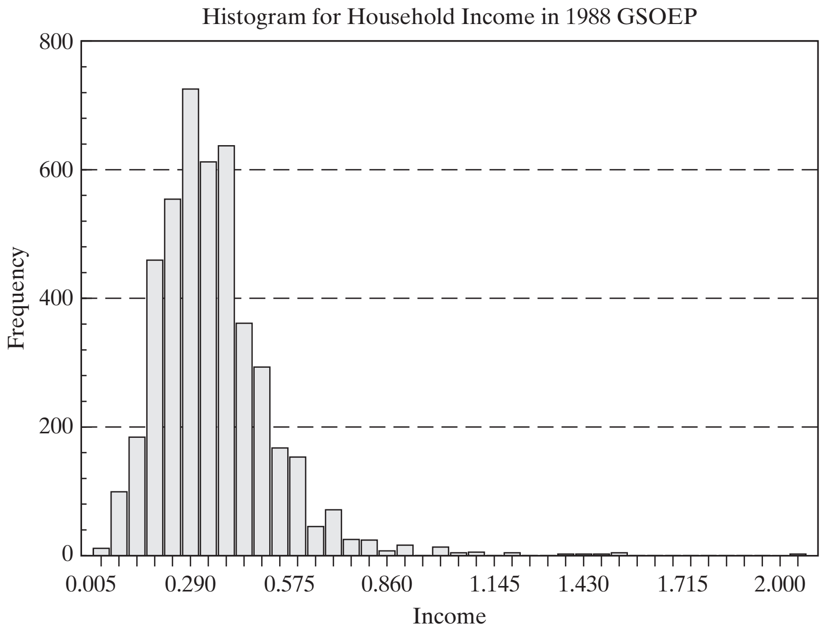

FIGURE 7.1 Histogram for Income.

This shows that an alternative way to handle the Box–Cox regression model is to transform the model into a nonlinear regression and then use the Gauss–Newton regression (see Section 7.2.6) to estimate the parameters. The original parameters of the model can be recovered by $\lambda = \gamma, \alpha = \alpha^* + 1/\gamma$ and $\beta = \gamma\beta^*$.

Example 7.6 *Interaction Effects in a Loglinear Model for Income*

A recent study in health economics is "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation" by Riphahn, Wambach, and Million (2003). The authors were interested in counts of physician visits and hospital visits and in the impact that the presence of private insurance had on the utilization counts of interest, that is, whether the data contain evidence of moral hazard. The sample used is an unbalanced panel of 7,293 households, the German Socioeconomic Panel (GSOEP) data set.[7] Among the variables reported in the panel are household income, with numerous other sociodemographic variables such as age, gender, and education. For this example, we will model the distribution of income using the last wave of the data set (1988), a cross section with 4,483 observations. Two of the individuals in this sample reported zero income, which is incompatible with the underlying models suggested in the development below. Deleting these two observations leaves a sample of 4,481 observations. Figures 7.1 and 7.2 display a histogram and a kernel density estimator for the household income variable for these observations.

We will fit an exponential regression model to the income variable, with

$$
\begin{aligned}
Income = \exp(&\beta_1 + \beta_2 Age + \beta_3 Age^2 + \beta_4 Education + \beta_5 Female \\
&+ \beta_6 Female \times Education + \beta_7 Age \times Education) + \varepsilon.
\end{aligned}
$$

[7]The data are published on the *Journal of Applied Econometrics* data archive Web site, at http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/. The variables in the data file are listed in Appendix Table F7.1. The number of observations in each year varies from one to seven with a total number of 27,326 observations. We will use these data in several examples here and later in the book.

(correctly) that the one defined in (13-4) would be optimal, once again based on the logic that motivates generalized least squares. This result is the now-celebrated one of Hansen (1982).

The asymptotic covariance matrix of this **generalized method of moments (GMM) estimator** is

$$\mathbf{V}_{GMM} = \frac{1}{n}[\boldsymbol{\Gamma}'\mathbf{W}\boldsymbol{\Gamma}]^{-1} = \frac{1}{n}[\boldsymbol{\Gamma}'\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}]^{-1}, \tag{13-5}$$

where $\boldsymbol{\Gamma}$ is the matrix of derivatives with jth row equal to

$$\boldsymbol{\Gamma}^j = \text{plim}\,\frac{\partial \bar{m}_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'},$$

and $\boldsymbol{\Phi} = \text{Asy. Var}[\sqrt{n}\,\bar{\mathbf{m}}]$. Finally, by virtue of the central limit theorem applied to the sample moments and the **Slutsky theorem** applied to this manipulation, we can expect the estimator to be asymptotically normally distributed. We will revisit the asymptotic properties of the estimator in Section 13.4.3.

Example 13.7 GMM Estimation of a Nonlinear Regression Model

In Example 7.6, we examined a nonlinear regression model for income using the German Socioeconomic Panel Data set. The regression model was

$$Income = h(1, Age, Education, Female, \boldsymbol{\gamma}) + \varepsilon,$$

where $h(.)$ is an exponential function of the variables. In the example, we used several interaction terms. In this application, we will simplify the conditional mean function somewhat, and use

$$Income = \exp(\gamma_1 + \gamma_2 Age + \gamma_3 Education + \gamma_4 Female) + \varepsilon,$$

which, for convenience, we will write

$$\begin{aligned} y_i &= \exp(\mathbf{x}_i'\boldsymbol{\gamma}) + \varepsilon_i \\ &= \mu_i + \varepsilon_i.^9 \end{aligned}$$

. The sample consists of the 1988 wave of the panel, less two observations for which *Income* equals zero. The resulting sample contains 4,481 observations. Descriptive statistics for the sample data are given in Table 7.2.

We ~~will first consider nonlinear least squares estimation of the parameters. The normal equations for nonlinear least squares will be~~

$$\cancel{(1/n)\,\Sigma_i[(y_i - \mu_i)\mu_i\mathbf{x}_i] = (1/n)\,\Sigma_i[\varepsilon_i\mu_i\mathbf{x}_i] = \mathbf{0}.}$$

~~Note that the orthogonality condition involves the pseudoregressors, $\partial\mu_i/\partial\boldsymbol{\gamma} = \mathbf{x}_i^0 = \mu_i\mathbf{x}_i$. The implied population moment equation is~~

$$\cancel{E[\varepsilon_i(\mu_i\mathbf{x}_i)] = \mathbf{0}.}$$

~~Computation of the nonlinear least squares estimator is discussed in Section 7.2.6. The estimator of the asymptotic covariance matrix is~~

$$\cancel{\text{Est. Asy. Var}[\hat{\boldsymbol{\gamma}}_{\text{NLSQ}}] = \frac{\Sigma_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{(4{,}481 - 4)}\left[\sum_{i=1}^{4{,}481}(\hat{\mu}_i\mathbf{x}_i)(\hat{\mu}_i\mathbf{x}_i)'\right]^{-1}, \text{ where } \hat{\mu}_i = \exp(\mathbf{x}_i'\hat{\boldsymbol{\gamma}}).}$$

[9] We note that in this model, it is likely that *Education* is endogenous. It would be straightforward to accommodate that in the GMM estimator. However, for purposes of a straightforward numerical example, we will proceed assuming that *Education* is exogenous

A simple method of moments estimator might be constructed from the hypothesis that $\mathbf{x}_i$ (not $\mathbf{x}_i^0$) is orthogonal to ε_i. Then,

$$E[\varepsilon_i \mathbf{x}_i] = E\left[\varepsilon_i \begin{pmatrix} 1 \\ Age_i \\ Education_i \\ Female_i \end{pmatrix}\right] = \mathbf{0}$$

implies four moment equations. The sample counterparts will be

$$\bar{m}_k(\gamma) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_i)x_{ik} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i x_{ik}.$$

In order to compute the method of moments estimator, we will minimize the sum of squares,

$$\bar{\mathbf{m}}'(\gamma)\bar{\mathbf{m}}(\gamma) = \sum_{k=1}^{4}\bar{m}_k^2(\gamma).$$

This is a nonlinear optimization problem that must be solved iteratively using the methods described in Section E.3.

With the first-step estimated parameters, $\hat{\gamma}^0$ in hand, the covariance matrix is estimated using (13-5).

$$\hat{\mathbf{\Phi}} = \left\{\frac{1}{4{,}481}\sum_{i=1}^{4{,}481}\mathbf{m}_i(\hat{\gamma}^0)\mathbf{m}_i'(\hat{\gamma}^0)\right\} = \left\{\frac{1}{4{,}481}\sum_{i=1}^{4{,}481}\left(\hat{\varepsilon}_i^0\mathbf{x}_i\right)\left(\hat{\varepsilon}_i^0\mathbf{x}_i\right)'\right\}$$

$$\bar{\mathbf{G}} = \left\{\frac{1}{4{,}481}\sum_{i=1}^{n}\left(\hat{\varepsilon}_i^0\mathbf{x}_i\right)\left(-\hat{\mu}_i^0\mathbf{x}_i\right)'\right\}.$$

The asymptotic covariance matrix for the MOM estimator is computed using (13-5),

$$\text{Est. Asy. Var}[\hat{\gamma}_{MOM}] = \frac{1}{n}[\bar{\mathbf{G}}\hat{\mathbf{\Phi}}^{-1}\bar{\mathbf{G}}']^{-1}.$$

Suppose we have in hand additional variables, *Health Satisfaction* and *Marital Status*, such that although the conditional mean function remains as given previously, we will use them to form a GMM estimator. This provides two additional moment equations,

$$E\left[\varepsilon_i\begin{pmatrix} Health\ Satisfaction_i \\ Marital\ Status_i \end{pmatrix}\right]$$

for a total of six moment equations for estimating the four parameters. We constuct the generalized method of moments estimator as follows: The initial step is the same as before, except the sum of squared moments, $\bar{\mathbf{m}}'(\gamma)\bar{\mathbf{m}}(\gamma)$, is summed over six rather than four terms. We then construct

$$\hat{\mathbf{\Phi}} = \left\{\frac{1}{4{,}481}\sum_{i=1}^{4{,}481}\mathbf{m}_i(\hat{\gamma})\,\mathbf{m}_i'(\hat{\gamma})\right\} = \left\{\frac{1}{4{,}481}\sum_{i=1}^{4{,}481}(\hat{\varepsilon}_i\mathbf{z}_i)(\hat{\varepsilon}_i\mathbf{z}_i)'\right\},$$

where now, $\mathbf{z}_i$ in the second term is the six exogenous variables, rather than the original four (including the constant term). Thus, $\hat{\mathbf{\Phi}}$ is now a 6×6 moment matrix. The optimal weighting matrix for estimation (developed in the next section) is $\hat{\mathbf{\Phi}}^{-1}$. The GMM estimator is computed by minimizing with respect to γ

$$q = \bar{\mathbf{m}}'(\gamma)\hat{\mathbf{\Phi}}^{-1}\bar{\mathbf{m}}(\gamma).$$

The asymptotic covariance matrix is computed using (13-5) as it was for the simple method of moments estimator.

TABLE 13.2 Nonlinear Regression Estimates (Standard Errors in Parentheses)

Estimate	~~*Nonlinear Least Squares*~~	*Method of Moments*	*First Step GMM*	*GMM*
Constant	~~−1.69331~~	−1.62969	−1.45551	−1.61192
	~~(0.04408)~~	(0.04214)	(0.10102)	(0.04163)
Age	~~0.00207~~	0.00178	−0.00028	0.00092
	~~(0.00061)~~	(0.00057)	(0.00100)	(0.00056)
Education	~~0.04792~~	0.04861	0.03731	0.04647
	~~(0.00247)~~	(0.00262)	(0.00518)	(0.00262)
Female	~~−0.00658~~	0.00070	−0.02205	−0.01517
	~~(0.01373)~~	(0.01384)	(0.01445)	(0.01357)

Table 13.2 presents four sets of estimates, nonlinear least squares, method of moments, first-step GMM, and GMM using the optimal weighting matrix. Two comparisons are noted. The method of moments produces slightly different results from the nonlinear least squares estimator. This is to be expected, since they are different criteria. Judging by the standard errors, the GMM estimator seems to provide a very slight improvement over the nonlinear least squares and method of moments estimators. The conclusion, though, would seem to be that the two additional moments (variables) do not provide very much additional information for estimation of the parameters.

13.4.3 PROPERTIES OF THE GMM ESTIMATOR

We will now examine the properties of the GMM estimator in some detail. Because the GMM estimator includes other familiar estimators that we have already encountered, including least squares (linear and nonlinear), and instrumental variables, these results will extend to those cases. The discussion given here will only sketch the elements of the formal proofs. The assumptions we make here are somewhat narrower than a fully general treatment might allow, but they are broad enough to include the situations likely to arise in practice. More detailed and rigorous treatments may be found in, for example, Newey and McFadden (1994), White (2001), Hayashi (2000), Mittelhammer et al. (2000), or Davidson (2000).

The GMM estimator is based on the set of population orthogonality conditions,

$$E[\mathbf{m}_i(\boldsymbol{\theta}_0)] = \mathbf{0},$$

where we denote the true parameter vector by $\boldsymbol{\theta}_0$. The subscript i on the term on the left-hand side indicates dependence on the observed data, $(\mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i)$. Averaging this over the sample observations produces the sample moment equation

$$E[\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0)] = \mathbf{0},$$

where

$$\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{m}_i(\boldsymbol{\theta}_0).$$

This moment is a set of L equations involving the K parameters. We will assume that this expectation exists and that the sample counterpart converges to it. The definitions are cast in terms of the population parameters and are indexed by the sample size. To fix the ideas, consider, once again, the empirical moment equations that define the instrumental variable estimator for a linear or nonlinear regression model.

Example 13.8 Empirical Moment Equation for Instrumental Variables
For the IV estimator in the linear or nonlinear regression model, we assume

$$E[\bar{\mathbf{m}}_n(\boldsymbol{\beta})] = E\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i[y_i - h(\mathbf{x}_i, \boldsymbol{\beta})]\right] = \mathbf{0}.$$

There are L instrumental variables in $\mathbf{z}_i$ and K parameters in $\boldsymbol{\beta}$. This statement defines L moment equations, one for each instrumental variable.

We make the following assumptions about the model and these empirical moments:

ASSUMPTION 13.1. **Convergence of the Empirical Moments:** *The data generating process is assumed to meet the conditions for a law of large numbers to apply, so that we may assume that the empirical moments converge in probability to their expectation. Appendix D lists several different laws of large numbers that increase in generality. What is required for this assumption is that*

$$\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i(\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{0}.$$

The laws of large numbers that we examined in Appendix D accommodate cases of independent observations. Cases of dependent or correlated observations can be gathered under the **Ergodic theorem** (20.1). For this more general case, then, we would assume that the sequence of observations $\mathbf{m}(\boldsymbol{\theta})$ constitutes a jointly $(L \times 1)$ stationary and ergodic process.

The empirical moments are assumed to be continuous and continuously differentiable functions of the parameters. For our earlier example, this would mean that the conditional mean function, $h(\mathbf{x}_i, \boldsymbol{\beta})$ is a continuous function of $\boldsymbol{\beta}$ (although not necessarily of $\mathbf{x}_i$). With continuity and differentiability, we will also be able to assume that the derivatives of the moments,

$$\bar{\mathbf{G}}_n(\boldsymbol{\theta}_0) = \frac{\partial \bar{\mathbf{m}}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathbf{m}_{i,n}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0'},$$

converge to a probability limit, say, plim $\bar{\mathbf{G}}_n(\boldsymbol{\theta}_0) = \bar{\mathbf{G}}(\boldsymbol{\theta}_0)$. [See (13-1), (13-5), and Theorem 13.1.] For sets of *independent* observations, the continuity of the functions and the derivatives will allow us to invoke the Slutsky theorem to obtain this result. For the more general case of sequences of *dependent* observations, Theorem 20.2, Ergodicity of Functions, will provide a counterpart to the Slutsky theorem for time-series data. In sum, if the moments themselves obey a law of large numbers, then it is reasonable to assume that the derivatives do as well.

ASSUMPTION 13.2. **Identification:** *For any $n \geq K$, if $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are two different parameter vectors, then there exist data sets such that $\bar{\mathbf{m}}_n(\boldsymbol{\theta}_1) \neq \bar{\mathbf{m}}_n(\boldsymbol{\theta}_2)$. Formally, in Section 12.5.3, identification is defined to imply that the probability limit of the GMM criterion function is uniquely minimized at the true parameters, $\boldsymbol{\theta}_0$.*

Assumption 13.2 is a practical prescription for identification. More formal conditions are discussed in Section 12.5.3. We have examined two violations of this crucial assumption. In the linear regression model, one of the assumptions is full rank of the matrix of exogenous variables—the absence of multicollinearity in $\mathbf{X}$. In our discussion of the maximum likelihood estimator, we will encounter a case (Example 14.1) in which a normalization is needed to identify the vector of parameters. [See Hansen et al. (1996) for discussion of this case.] Both of these cases are included in this assumption. The identification condition has three important implications:

1. **Order condition.** The number of moment conditions is at least as large as the number of parameters; $L \geq K$. This is necessary, but not sufficient for identification.
2. **Rank condition.** The $L \times K$ matrix of derivatives, $\bar{\mathbf{G}}_n(\boldsymbol{\theta}_0)$ will have row rank equal to K. (Again, note that the number of rows must equal or exceed the number of columns.)
3. **Uniqueness.** With the continuity assumption, the identification assumption implies that the parameter vector that satisfies the population moment condition is unique. We know that at the true parameter vector, plim $\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) = \mathbf{0}$. If $\boldsymbol{\theta}_1$ is any parameter vector that satisfies this condition, then $\boldsymbol{\theta}_1$ must equal $\boldsymbol{\theta}_0$.

Assumptions 13.1 and 13.2 characterize the parameterization of the model. Together they establish that the parameter vector will be estimable. We now make the statistical assumption that will allow us to establish the properties of the GMM estimator.

ASSUMPTION 13.3. **Asymptotic Distribution of Empirical Moments:** *We assume that the empirical moments obey a central limit theorem. This assumes that the moments have a finite asymptotic covariance matrix,* $(1/n)\boldsymbol{\Phi}$*, so that*

$$\sqrt{n}\,\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) \xrightarrow{d} N[\mathbf{0}, \boldsymbol{\Phi}].$$

The underlying requirements on the data for this assumption to hold will vary and will be complicated if the observations comprising the empirical moment are not independent. For samples of independent observations, we assume the conditions underlying the Lindeberg–Feller (D.19) or Liapounov central limit theorem (D.20) will suffice. For the more general case, it is once again necessary to make some assumptions about the data. We have assumed that

$$E[\mathbf{m}_i(\boldsymbol{\theta}_0)] = \mathbf{0}.$$

If we can go a step further and assume that the functions $\mathbf{m}_i(\boldsymbol{\theta}_0)$ are an ergodic, stationary **martingale difference series**,

$$E[\mathbf{m}_i(\boldsymbol{\theta}_0) \mid \mathbf{m}_{i-1}(\boldsymbol{\theta}_0), \mathbf{m}_{i-2}(\boldsymbol{\theta}_0) \ldots] = \mathbf{0},$$

then we can invoke Theorem 20.3, the central limit theorem for the Martingale difference series. It will generally be fairly complicated to verify this assumption for nonlinear models, so it will usually be assumed outright. On the other hand, the assumptions are likely to be fairly benign in a typical application. For regression models, the assumption takes the form

$$E[\mathbf{z}_i\varepsilon_i \mid \mathbf{z}_{i-1}\varepsilon_{i-1}, \ldots] = \mathbf{0},$$

which will often be part of the central structure of the model.

With the assumptions in place, we have

THEOREM 13.2 Asymptotic Distribution of the GMM Estimator
Under the preceding assumptions,

$$\hat{\boldsymbol{\theta}}_{GMM} \xrightarrow{p} \boldsymbol{\theta}_0,$$
$$\hat{\boldsymbol{\theta}}_{GMM} \overset{a}{\sim} N[\boldsymbol{\theta}_0, \mathbf{V}_{GMM}], \tag{13-6}$$

where $\mathbf{V}_{GMM}$ is defined in (13-5).

We will now sketch a proof of Theorem 13.2. The GMM estimator is obtained by minimizing the criterion function

$$q_n(\boldsymbol{\theta}) = \bar{\mathbf{m}}_n(\boldsymbol{\theta})'\mathbf{W}_n\bar{\mathbf{m}}_n(\boldsymbol{\theta}),$$

where $\mathbf{W}_n$ is the weighting matrix used. Consistency of the estimator that minimizes this criterion can be established by the same logic that will be used for the maximum likelihood estimator. It must first be established that $q_n(\boldsymbol{\theta})$ converges to a value $q_0(\boldsymbol{\theta})$. By our assumptions of strict continuity and Assumption 13.1, $q_n(\boldsymbol{\theta}_0)$ converges to 0. (We could apply the Slutsky theorem to obtain this result.) We will assume that $q_n(\boldsymbol{\theta})$ converges to $q_0(\boldsymbol{\theta})$ for other points in the parameter space as well. Because $\mathbf{W}_n$ is positive definite, for any finite n, we know that

$$0 \leq q_n(\hat{\boldsymbol{\theta}}_{GMM}) \leq q_n(\boldsymbol{\theta}_0). \tag{13-7}$$

That is, in the finite sample, $\hat{\boldsymbol{\theta}}_{GMM}$ actually minimizes the function, so the sample value of the criterion is not larger at $\hat{\boldsymbol{\theta}}_{GMM}$ than at any other value, including the true parameters. But, at the true parameter values, $q_n(\boldsymbol{\theta}_0) \xrightarrow{p} 0$. So, if (13-7) is true, then it must follow that $q_n(\hat{\boldsymbol{\theta}}_{GMM}) \xrightarrow{p} \boldsymbol{\theta}_0$ as well because of the identification assumption, 13.2. As $n \to \infty$, $q_n(\hat{\boldsymbol{\theta}}_{GMM})$ and $q_n(\boldsymbol{\theta})$ converge to the same limit. It must be the case, then, that as $n \to \infty$, $\bar{\mathbf{m}}_n(\hat{\boldsymbol{\theta}}_{GMM}) \to \bar{\mathbf{m}}_n(\boldsymbol{\theta}_0)$, because the function is quadratic and $\mathbf{W}$ is positive definite. The identification condition that we assumed earlier now assures that as $n \to \infty$, $\hat{\boldsymbol{\theta}}_{GMM}$ must equal $\boldsymbol{\theta}_0$. This establishes consistency of the estimator.

We will now sketch a proof of the asymptotic normality of the estimator: The first-order conditions for the GMM estimator are

$$\frac{\partial q_n(\hat{\boldsymbol{\theta}}_{GMM})}{\partial \hat{\boldsymbol{\theta}}_{GMM}} = 2\bar{\mathbf{G}}_n(\hat{\boldsymbol{\theta}}_{GMM})'\mathbf{W}_n\bar{\mathbf{m}}_n(\hat{\boldsymbol{\theta}}_{GMM}) = \mathbf{0}. \tag{13-8}$$

(The leading 2 is irrelevant to the solution, so it will be dropped at this point.) The orthogonality equations are assumed to be continuous and continuously differentiable. This allows us to employ the **mean value theorem** as we expand the empirical moments in a linear Taylor series around the true value, $\boldsymbol{\theta}_0$

$$\bar{\mathbf{m}}_n(\hat{\boldsymbol{\theta}}_{GMM}) = \bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) + \bar{\mathbf{G}}_n(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0), \tag{13-9}$$

where $\bar{\boldsymbol{\theta}}$ is a point between $\hat{\boldsymbol{\theta}}_{GMM}$ and the true parameters, $\boldsymbol{\theta}_0$. Thus, for each element $\bar{\theta}_k = w_k\hat{\theta}_{k,GMM} + (1 - w_k)\theta_{0,k}$ for some w_k such that $0 < w_k < 1$. Insert (13-9) in (13-8) to obtain

$$\bar{\mathbf{G}}_n(\hat{\boldsymbol{\theta}}_{GMM})'\mathbf{W}_n\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) + \bar{\mathbf{G}}_n(\hat{\boldsymbol{\theta}}_{GMM})'\mathbf{W}_n\bar{\mathbf{G}}_n(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0) = \mathbf{0}.$$

Solve this equation for the estimation error and multiply by $\sqrt{n}$. This produces

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0) = -[\bar{\mathbf{G}}_n(\hat{\boldsymbol{\theta}}_{GMM})'\mathbf{W}_n\bar{\mathbf{G}}_n(\bar{\boldsymbol{\theta}})]^{-1}\bar{\mathbf{G}}_n(\hat{\boldsymbol{\theta}}_{GMM})'\mathbf{W}_n\sqrt{n}\,\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0).$$

Assuming that they have them, the quantities on the left- and right-hand sides have the same limiting distributions. By the consistency of $\hat{\boldsymbol{\theta}}_{GMM}$, we know that $\hat{\boldsymbol{\theta}}_{GMM}$ and $\bar{\boldsymbol{\theta}}$ both converge to $\boldsymbol{\theta}_0$. By the strict continuity assumed, it must also be the case that

$$\bar{\mathbf{G}}_n(\bar{\boldsymbol{\theta}}) \xrightarrow{p} \bar{\mathbf{G}}(\boldsymbol{\theta}_0) \text{ and } \bar{\mathbf{G}}_n(\hat{\boldsymbol{\theta}}_{GMM}) \xrightarrow{p} \bar{\mathbf{G}}(\boldsymbol{\theta}_0).$$

We have also assumed that the weighting matrix, $\mathbf{W}_n$, converges to a matrix of constants, $\mathbf{W}$. Collecting terms, we find that the limiting distribution of the vector on the left-hand side must be the same as that on the right-hand side in (13-10),

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0) \xrightarrow{d} \left\{-[\bar{\mathbf{G}}(\boldsymbol{\theta}_0)'\mathbf{W}\bar{\mathbf{G}}(\boldsymbol{\theta}_0)]^{-1}\bar{\mathbf{G}}(\boldsymbol{\theta}_0)'\mathbf{W}\right\}\sqrt{n}\,\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0). \tag{13-10}$$

We now invoke Assumption 13.3. The matrix in curled brackets is a set of constants. The last term has the normal limiting distribution given in Assumption 13.3. The mean and variance of this limiting distribution are zero and $\boldsymbol{\Phi}$, respectively. Collecting terms, we have the result in Theorem 13.2, where

$$\mathbf{V}_{GMM} = \frac{1}{n}[\bar{\mathbf{G}}(\boldsymbol{\theta}_0)'\mathbf{W}\bar{\mathbf{G}}(\boldsymbol{\theta}_0)]^{-1}\bar{\mathbf{G}}(\boldsymbol{\theta}_0)'\mathbf{W}\boldsymbol{\Phi}\mathbf{W}\bar{\mathbf{G}}(\boldsymbol{\theta}_0)[\bar{\mathbf{G}}(\boldsymbol{\theta}_0)'\mathbf{W}\bar{\mathbf{G}}(\boldsymbol{\theta}_0)]^{-1}. \tag{13-11}$$

The final result is a function of the choice of weighting matrix, $\mathbf{W}$. If the optimal weighting matrix, $\mathbf{W} = \boldsymbol{\Phi}^{-1}$, is used, then the expression collapses to

$$\mathbf{V}_{GMM,optimal} = \frac{1}{n}[\bar{\mathbf{G}}(\boldsymbol{\theta}_0)'\boldsymbol{\Phi}^{-1}\bar{\mathbf{G}}(\boldsymbol{\theta}_0)]^{-1}. \tag{13-12}$$

Returning to (13-11), there is a special case of interest. If we use least squares or instrumental variables with $\mathbf{W} = \mathbf{I}$, then

$$\mathbf{V}_{GMM} = \frac{1}{n}(\bar{\mathbf{G}}'\bar{\mathbf{G}})^{-1}\bar{\mathbf{G}}'\boldsymbol{\Phi}\bar{\mathbf{G}}(\bar{\mathbf{G}}'\bar{\mathbf{G}})^{-1}.$$

This equation prescibes essentially the White or **Newey-West estimator**, which returns us to our departure point and provides a neat symmetry to the GMM principle. We will formalize this in Section 13.6.1.