The Wald statistic can be computed using only the unrestricted estimate. The LM statistic is

$$\text{LM} = \text{Lagrange multiplier} = \mathbf{g}_1'(\mathbf{c}_0)\{\text{Est. Asy. Var}[\mathbf{g}_1(\mathbf{c}_0)]\}^{-1}\mathbf{g}_1(\mathbf{c}_0), \qquad (15\text{-}16)$$

where

$$\mathbf{g}_1(\mathbf{c}_0) = \partial \ln L_1(\mathbf{c}_0)/\partial \mathbf{c}_0,$$

that is, the first derivatives of the *unconstrained* log-likelihood computed at the *restricted* estimates. The term Est. Asy. Var$[\mathbf{g}_1(\mathbf{c}_0)]$ is the inverse of any of the usual estimators of the asymptotic covariance matrix of the maximum likelihood estimators of the parameters, computed using the restricted estimates. The most convenient choice is usually the BHHH estimator. The LM statistic is based on the restricted estimates.

Newey and West (1987b) have devised counterparts to these test statistics for the GMM estimator. The Wald statistic is computed identically, using the results of GMM estimation rather than maximum likelihood.[10] That is, in (15-14), we would use the unrestricted GMM estimator of $\boldsymbol{\theta}$. The appropriate asymptotic covariance matrix is (15-12). The computation is exactly the same. The counterpart to the LR statistic is the difference in the values of nq in (15-13). It is necessary to use the same weighting matrix, $\mathbf{W}$, in both restricted and unrestricted estimators. Because the unrestricted estimator is consistent under both H_0 and H_1, a consistent, unrestricted estimator of $\boldsymbol{\theta}$ is used to compute $\mathbf{W}$. Label this $\boldsymbol{\Phi}_1^{-1} = \{\text{Asy. Var}[\sqrt{n}\,\overline{\mathbf{m}}_1(\mathbf{c}_1)]\}^{-1}$. In each occurrence, the subscript 1 indicates reference to the unrestricted estimator. Then q is minimized without restrictions to obtain q_1 and then subject to the restrictions to obtain q_0. The statistic is then $(nq_0 - nq_1)$.[11] Because we are using the same $\mathbf{W}$ in both cases, this statistic is necessarily nonnegative. (This is the statistic discussed in Section 15.5.1.)

Finally, the counterpart to the LM statistic would be

$$\text{LM}_{GMM} = n\left[\overline{\mathbf{m}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{G}}_1(\mathbf{c}_0)\right]\left[\overline{\mathbf{G}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{G}}_1(\mathbf{c}_0)\right]^{-1}\left[\overline{\mathbf{G}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{m}}_1(\mathbf{c}_0)\right].$$

The logic for this LM statistic is the same as that for the MLE. The derivatives of the minimized criterion q in (15-3) evaluated at the restricted estimator are

$$\mathbf{g}_1(\mathbf{c}_0) = \frac{\partial q}{\partial \mathbf{c}_0} = 2\overline{\mathbf{G}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{m}}(\mathbf{c}_0).$$

The **LM statistic,** LM_{GMM}, is a Wald statistic for testing the hypothesis that this vector equals zero under the restrictions of the null hypothesis. From our earlier results, we would have

$$\text{Est. Asy. Var}[\mathbf{g}_1(\mathbf{c}_0)] = \frac{4}{n}\overline{\mathbf{G}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\{\text{Est. Asy. Var}[\sqrt{n}\,\overline{\mathbf{m}}(\mathbf{c}_0)]\}\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{G}}_1(\mathbf{c}_0).$$

The estimated asymptotic variance of $\sqrt{n}\,\overline{\mathbf{m}}(\mathbf{c}_0)$ is $\hat{\boldsymbol{\Phi}}_1$, so

$$\text{Est. Asy. Var}[\mathbf{g}_1(\mathbf{c}_0)] = \frac{4}{n}\overline{\mathbf{G}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{G}}_1(\mathbf{c}_0).$$

[10] See Burnside and Eichenbaum (1996) for some small-sample results on this procedure. Newey and McFadden (1994) have shown the asymptotic equivalence of the three procedures.

[11] Newey and West label this test the D test.

13-31

The Wald statistic would be

$$\begin{aligned}\text{Wald} &= \mathbf{g}_1(\mathbf{c}_0)'\{\text{Est. Asy. Var}[\mathbf{g}_1(\mathbf{c}_0)]\}^{-1}\mathbf{g}_1(\mathbf{c}_0)\\ &= n\,\overline{\mathbf{m}}_1'(\mathbf{c}_0)\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{G}}_1(\mathbf{c}_0)\{\overline{\mathbf{G}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{G}}_1(\mathbf{c}_0)\}^{-1}\overline{\mathbf{G}}_1(\mathbf{c}_0)'\hat{\boldsymbol{\Phi}}_1^{-1}\overline{\mathbf{m}}_1(\mathbf{c}_0).\end{aligned} \qquad (~~15~~ 13\text{-}17)$$

~~15~~ 13.6 GMM ESTIMATION OF ECONOMETRIC MODELS

The preceding has suggested that the GMM approach to estimation broadly encompasses most of the estimators we will encounter in this book. We have implicitly examined least squares and the general method of instrumental variables in the process. In this section, we will formalize more specifically the GMM estimators for several of the estimators that appear in the earlier chapters. Section ~~15~~ 13.6.1 examines the generalized regression model of Chapter ~~8~~ 9. Section ~~15~~ 13.6.2 describes a relatively minor extension of the GMM/IV estimator to nonlinear regressions. Sections ~~15~~ 13.6.3 and ~~15~~ 13.6.4 describe the GMM estimators for our models of systems of equations, the seemingly unrelated regressions (SUR) model and models of simultaneous equations. In the latter, as we did in Chapter ~~13~~ 10, we consider both limited (single-equation) and full information (multiple-equation) estimators. Finally, in Section ~~15~~ 13.6.5, we develop one of the major applications of GMM estimation, the Arellano–Bond–Bover estimator for dynamic panel data models.

~~15~~ 13.6.1 SINGLE-EQUATION LINEAR MODELS

It is useful to confine attention to the instrumental variables case, as it is fairly general and we can easily specialize it to the simpler regression models if that is appropriate. Thus, we depart from the usual linear model (~~12-1~~ 8-1), but we no longer require that $E[\varepsilon_i \,|\, \mathbf{x}_i] = 0$. Instead, we adopt the instrumental variables formulation in ~~Section 12.6~~ Chapter 8. That is, our model is

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$
$$E[\mathbf{z}_i\varepsilon_i] = \mathbf{0}$$

for K variables in $\mathbf{x}_i$ and for some set of L instrumental variables, $\mathbf{z}_i$, where $L \geq K$. The earlier case of the generalized regression model arises if $\mathbf{z}_i = \mathbf{x}_i$, and the classical regression form results if we add $\boldsymbol{\Omega} = \mathbf{I}$ as well, so this is a convenient encompassing model framework.

In Chapter ~~8~~ 9 on generalized least squares estimation, we considered two cases, first one with a known $\boldsymbol{\Omega}$, then one with an unknown $\boldsymbol{\Omega}$ that must be estimated. In estimation by the generalized method of moments, neither of these approaches is relevant because we begin with much less (assumed) knowledge about the data generating process. We will consider three cases:

- Classical regression: $\text{Var}[\varepsilon_i \,|\, \mathbf{X}, \mathbf{Z}] = \sigma^2$,
- Heteroscedasticity: $\text{Var}[\varepsilon_i \,|\, \mathbf{X}, \mathbf{Z}] = \sigma_i^2$,
- Generalized model: $\text{Cov}[\varepsilon_t, \varepsilon_s \,|\, \mathbf{X}, \mathbf{Z}] = \sigma^2\omega_{ts}$,

13-32

where $\mathbf{Z}$ and $\mathbf{X}$ are the $n \times L$ and $n \times K$ observed data matrices. (We assume, as will often be true, that the fully general case will apply in a time-series setting. Hence the change in the subscripts.) *No specific distribution is assumed for the disturbances, conditional or unconditional.*

KT

The assumption $E[\mathbf{z}_i \varepsilon_i] = \mathbf{0}$ implies the following **orthogonality condition:**

$$\text{Cov}[\mathbf{z}_i, \varepsilon_i] = \mathbf{0}, \quad \text{or} \quad E[\mathbf{z}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})] = \mathbf{0}.$$

KT

By summing the terms, we find that this further implies the **population moment equation,**

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})\right] = E[\overline{\mathbf{m}}(\boldsymbol{\beta})] = \mathbf{0}. \qquad 13\ (\cancel{15}\text{-}18)$$

This relationship suggests how we might now proceed to estimate $\boldsymbol{\beta}$. Note, in fact, that if $\mathbf{z}_i = \mathbf{x}_i$, then this is just the population counterpart to the least squares normal equations. So, as a guide to estimation, this would return us to least squares. Suppose, we now translate this population expectation into a sample analog and use that as our guide for estimation. That is, if the population relationship holds for the true parameter vector, $\boldsymbol{\beta}$, suppose we attempt to mimic this result with a sample counterpart, or empirical moment equation,

$$\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})\right] = \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i(\hat{\boldsymbol{\beta}})\right] = \overline{\mathbf{m}}(\hat{\boldsymbol{\beta}}) = \mathbf{0}. \qquad 13\ (\cancel{15}\text{-}19)$$

In the absence of other information about the data generating process, we can use the empirical moment equation as the basis of our estimation strategy.

The empirical moment condition is L equations (the number of variables in $\mathbf{Z}$) in K unknowns (the number of parameters we seek to estimate). There are three possibilities to consider:

1. **Underidentified:** $L < K$. If there are fewer moment equations than there are parameters, then it will not be possible to find a solution to the equation system in (~~15~~ 13-19). With no other information, such as restrictions that would reduce the number of free parameters, there is no need to proceed any further with this case.

For the identified cases, it is convenient to write (~~15~~ 13-19) as

$$\overline{\mathbf{m}}(\hat{\boldsymbol{\beta}}) = \left(\frac{1}{n}\mathbf{Z}'\mathbf{y}\right) - \left(\frac{1}{n}\mathbf{Z}'\mathbf{X}\right)\hat{\boldsymbol{\beta}}. \qquad 13\ (\cancel{15}\text{-}20)$$

2. **Exactly identified.** If $L = K$, then you can easily show (we leave it as an exercise) that the single solution to our equation system is the familiar instrumental variables estimator from Section 12.3.2,

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}. \qquad 13\ (\cancel{15}\text{-}21)$$

3. **Overidentified.** If $L > K$, then there is no unique solution to the equation system $\overline{\mathbf{m}}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$. In this instance, we need to formulate some strategy to choose an estimator. One intuitively appealing possibility which has served well thus far is "least squares." In this instance, that would mean choosing the estimator based on the criterion function

$$\text{Min}_{\boldsymbol{\beta}}\ q = \overline{\mathbf{m}}(\hat{\boldsymbol{\beta}})'\overline{\mathbf{m}}(\hat{\boldsymbol{\beta}}).$$

We do keep in mind that we ~~will~~ only be able to minimize this at some positive value; there is no exact solution to (~~15~~13-19) in the overidentified case. Also, you can verify that if we treat the exactly identified case as if it were overidentified, that is, use least squares anyway, we will still obtain the IV estimator shown in (~~15~~13-21) for the solution to case (2). For the overidentified case, the first-order conditions are

$$\frac{\partial q}{\partial \beta} = 2\left(\frac{\partial \bar{\mathbf{m}}'(\hat{\beta})}{\partial \beta}\right)\bar{\mathbf{m}}(\hat{\beta}) = 2\bar{\mathbf{G}}(\hat{\beta})'\bar{\mathbf{m}}(\hat{\beta})$$
$$= 2\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}\right) = \mathbf{0}. \qquad (\cancel{15}\,13\text{-}22)$$

We leave as exercise to show that the solution in both cases (2) and (3) is now

$$\hat{\beta} = [(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{y}). \qquad (\cancel{15}\,13\text{-}23)$$

The estimator in (~~15~~13-23) is a hybrid that we have ~~not~~ encountered before, though if $L = K$, then it does reduce to the earlier one in (~~15~~13-21). (In the overidentified case, (~~15~~13-21) is not an IV estimator, it is, as we have sought, a **method of moments estimator.**)

AU: Term is already a KT on ms p 13-4. Here also?

It remains to establish consistency and to obtain the asymptotic distribution and an asymptotic covariance matrix for the estimator. The intermediate results we need are Assumptions ~~15~~13.1, ~~15~~13.2 and ~~15~~13.3 in Section ~~15~~13.4.3:

- **Convergence of the moments.** The sample moment converges in probability to its population counterpart. That is, $\bar{\mathbf{m}}(\beta) \to \mathbf{0}$. Different circumstances will produce different kinds of convergence, but we will require it in some form. For the simplest cases, such as a model of heteroscedasticity, this will be convergence in mean square. Certain time-series models that involve correlated observations will necessitate some other form of convergence. But, in any of the cases we consider, we will require the general result: plim $\bar{\mathbf{m}}(\beta) = \mathbf{0}$.
- **Identification.** The parameters are identified in terms of the moment equations. Identification means, essentially, that a large enough sample will contain sufficient information for us actually to estimate β consistently using the sample moments. KT There are two conditions which must be met—an **order condition,** which we have already assumed ($L \geq K$), and a **rank condition,** which states that the moment equations are not redundant. The rank condition implies the order condition, so we need only formalize it: KT
- **Identification condition for GMM estimation:** The $L \times K$ matrix

$$\Gamma(\beta) = E[\bar{\mathbf{G}}(\beta)] = \text{plim}\ \bar{\mathbf{G}}(\beta) = \text{plim}\frac{\partial \bar{\mathbf{m}}}{\partial \beta'} = \text{plim}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathbf{m}_i}{\partial \beta'}$$

FN 12

must have row rank equal to K.[12] Because this requires $L \geq K$, this implies the order condition. This assumption means that this derivative matrix converges in probability to its expectation. Note that we have assumed, in addition, that the

[12]We require that the row rank be at least as large as K. There could be redundant, that is, functionally dependent, moments, so long as there are at least K that are functionally independent.

13-34

derivatives, like the moments themselves, obey a law of large numbers—they converge in probability to their expectations.

- **Limiting Normal Distribution for the Sample Moments.** The population moment obeys a central limit theorem or some similar variant. Since we are studying a generalized regression model, Lindeberg–Levy (D.18.) will be too narrow—the observations will have different variances. Lindeberg–Feller (D.19.A) suffices in the heteroscedasticity case, but in the general case, we will ultimately require something more general. See Section ~~12~~13.4.3.

It will follow from Assumptions ~~15~~13.1–~~15~~13.3 (again, at this point we do this without proof) that the GMM estimators that we obtain are, in fact, consistent. By virtue of the Slutsky theorem, we can transfer our limiting results to the empirical moment equations.

To obtain the asymptotic covariance matrix we will simply invoke the general result for GMM estimators in Section ~~15~~13.4.3. That is,

$$\text{Asy. Var}[\hat{\beta}] = \frac{1}{n}[\boldsymbol{\Gamma}'\boldsymbol{\Gamma}]^{-1}\boldsymbol{\Gamma}'\{\text{Asy. Var}[\sqrt{n}\,\overline{\mathbf{m}}(\boldsymbol{\beta})]\}\boldsymbol{\Gamma}[\boldsymbol{\Gamma}'\boldsymbol{\Gamma}]^{-1}.$$

For the particular model we are studying here,

$$\overline{\mathbf{m}}(\boldsymbol{\beta}) = (1/n)(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\boldsymbol{\beta}),$$
$$\overline{\mathbf{G}}(\boldsymbol{\beta}) = (1/n)\mathbf{Z}'\mathbf{X},$$
$$\boldsymbol{\Gamma}(\boldsymbol{\beta}) = \mathbf{Q}_{ZX} \text{ (see Section } \cancel{12}8.3.2).$$

(You should check in the preceding expression that the dimensions of the particular matrices and the dimensions of the various products produce the correctly configured matrix that we seek.) The remaining detail, which is the crucial one for the model we are examining, is for us to determine

$$\mathbf{V} = \text{Asy. Var}[\sqrt{n}\,\overline{\mathbf{m}}(\boldsymbol{\beta})].$$

Given the form of $\overline{\mathbf{m}}(\boldsymbol{\beta})$,

$$\mathbf{V} = \frac{1}{n}\text{Var}\left[\sum_{i=1}^{n}\mathbf{z}_i\varepsilon_i\right] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma^2\omega_{ij}\mathbf{z}_i\mathbf{z}_j' = \sigma^2\frac{\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}}{n}$$

for the most general case. Note that this is precisely the expression that appears in (8-6), so the question that arose there arises here once again. That is, under what conditions will this converge to a constant matrix? We take the discussion there as given. The only remaining detail is how to estimate this matrix. The answer appears in Section 8.2.3, where we pursued this same question in connection with robust estimation of the asymptotic covariance matrix of the least squares estimator. To review then, what we have achieved to this point is to provide a theoretical foundation for the instrumental variables estimator. As noted earlier, this specializes to the least squares estimator. The estimators of V for our three cases will be

- Classical regression:

$$\hat{\mathbf{V}} = \frac{(\mathbf{e}'\mathbf{e}/n)}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i' = \frac{(\mathbf{e}'\mathbf{e}/n)}{n}\mathbf{Z}'\mathbf{Z}.$$

- Heteroscedastic regression:

$$\hat{\mathbf{V}} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{z}_i \mathbf{z}_i'. \qquad (\cancel{15}\,13\text{-}24)$$

- Generalized regression:

$$\hat{\mathbf{V}} = \frac{1}{n}\left[\sum_{t=1}^{n} e_t^2 \mathbf{z}_t \mathbf{z}_t' + \sum_{\ell=1}^{p}\left(1 - \frac{\ell}{(p+1)}\right)\sum_{t=\ell+1}^{n} e_t e_{t-\ell}(\mathbf{z}_t \mathbf{z}_{t-\ell}' + \mathbf{z}_{t-\ell}\mathbf{z}_t')\right].$$

We should observe that in each of these cases, we have actually used some information about the structure of $\boldsymbol{\Omega}$. If it is known only that the terms in $\bar{\mathbf{m}}(\beta)$ are uncorrelated, then there is a convenient estimator available,

$$\hat{\mathbf{V}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{m}_i(\hat{\beta})\mathbf{m}_i(\hat{\beta})',$$

that is, the natural, empirical variance estimator. Note that this is what is being used in the heteroscedasticity case directly preceding.

Collecting all the terms so far, then, we have

$$\begin{aligned} \text{Est. Asy. Var}[\hat{\beta}] &= \frac{1}{n}[\bar{\mathbf{G}}(\hat{\beta})'\bar{\mathbf{G}}(\hat{\beta})]^{-1}\bar{\mathbf{G}}(\hat{\beta})'\hat{\mathbf{V}}\bar{\mathbf{G}}(\hat{\beta})[\bar{\mathbf{G}}(\hat{\beta})'\bar{\mathbf{G}}(\hat{\beta})]^{-1} \\ &= n[(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{Z})\hat{\mathbf{V}}(\mathbf{Z}'\mathbf{X})[(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{X})]^{-1}. \end{aligned} \qquad (\cancel{15}\,13\text{-}25)$$

The preceding might seem to endow the least squares or method of moments estimators with some degree of optimality, but that is not the case. We have only provided them with a different statistical motivation (and established consistency). We now consider the question of whether, because this is the generalized regression model, there is some better (more efficient) means of using the data.

The class of minimum distance estimators for this model is defined by the solutions to the criterion function

$$\text{Min}_{\beta}\, q = \bar{\mathbf{m}}(\beta)'\mathbf{W}\bar{\mathbf{m}}(\beta),$$

where $\mathbf{W}$ is *any* positive definite **weighting matrix.** Based on the assumptions just made, we can invoke Theorem ~~15~~13.1 to obtain

AA

$$\text{Asy. Var}\left[\hat{\beta}_{MD}\right] = \frac{1}{n}\left[\bar{\mathbf{G}}'\mathbf{W}\bar{\mathbf{G}}\right]^{-1}\bar{\mathbf{G}}'\mathbf{W}\mathbf{V}\mathbf{W}\bar{\mathbf{G}}\left[\bar{\mathbf{G}}'\mathbf{W}\bar{\mathbf{G}}\right]^{-1}.$$

AU: Term "weighting matrix" a KT on msp 13-12. Here also?

Note that our entire preceding analysis was of the simplest minimum distance estimator, which has $\mathbf{W} = \mathbf{I}$. The obvious question now arises, if any $\mathbf{W}$ produces a consistent estimator, is any $\mathbf{W}$ better than any other one, or is it simply arbitrary? There is a firm answer, for which we have to consider two cases separately:

- **Exactly identified case.** If $L = K$; that is, if the number of moment conditions is the same as the number of parameters being estimated, then $\mathbf{W}$ is irrelevant to the solution, so on the basis of simplicity alone, the optimal $\mathbf{W}$ is $\mathbf{I}$.
- **Overidentified case.** In this case, the "optimal" weighting matrix, that is, the $\mathbf{W}$ that produces the most efficient estimator, is $\mathbf{W} = \mathbf{V}^{-1}$. The best weighting matrix is the inverse of the asymptotic covariance of the moment vector. In this case,

13-36

the MDE will be the GMM estimator with

$$\hat{\beta}_{GMM} = [(\mathbf{X}'\mathbf{Z})\hat{\mathbf{V}}^{-1}(\mathbf{Z}'\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{Z})\hat{\mathbf{V}}^{-1}(\mathbf{Z}'\mathbf{y}),$$

and

$$\text{Asy. Var}[\hat{\beta}_{GMM}] = \frac{1}{n}[\bar{\mathbf{G}}'\mathbf{V}^{-1}\bar{\mathbf{G}}]^{-1}$$
$$= [(\mathbf{X}'\mathbf{Z})\mathbf{V}^{-1}(\mathbf{Z}'\mathbf{X})]^{-1}.$$

n

We conclude this discussion by tying together what should seem to be a loose end. The GMM estimator is computed as the solution to

$$\text{Min}_{\beta}\ q = \bar{\mathbf{m}}(\beta)'\left\{\text{Asy. Var}[\sqrt{n}\,\bar{\mathbf{m}}(\beta)]\right\}^{-1}\bar{\mathbf{m}}(\beta),$$

which might suggest that the weighting matrix is a function of the thing we are trying to estimate. The process of GMM estimation will have to proceed in two steps: Step 1 is to obtain an estimate of $\mathbf{V}$; Step 2 will consist of using the inverse of this $\mathbf{V}$ as the weighting matrix in computing the GMM estimator. The following is a common strategy:

Step 1. Use $\mathbf{W} = \mathbf{I}$ to obtain a consistent estimator of β. Then, estimate $\mathbf{V}$ with

$$\hat{\mathbf{V}} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{z}_i \mathbf{z}_i'$$

in the heteroscedasticity case (i.e., the White estimator) or, for the more general case, the Newey-West estimator.

Step 2. Use $\mathbf{W} = \hat{\mathbf{V}}^{-1}$ to compute the GMM estimator.

By this point, the observant reader should have noticed that in all of the preceding, we have never actually encountered the two-stage least squares estimator that we introduced in Section 12.3.3. To obtain this estimator, we must revert back to the classical, that is, homoscedastic, and nonautocorrelated disturbances case. In that instance, the weighting matrix in Theorem 15.2 will be $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ and we will obtain the apparently missing result.

13c / AU: Term "GMM estimator" is KT on msp 13-19, Here also?

The **GMM estimator** in the heteroscedastic regression model is produced by the empirical moment equations

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\left(y_i - \mathbf{x}_i'\hat{\beta}_{GMM}\right) = \frac{1}{n}\mathbf{X}'\hat{\boldsymbol{\varepsilon}}\left(\hat{\beta}_{GMM}\right) = \bar{\mathbf{m}}\left(\hat{\beta}_{GMM}\right) = \mathbf{0}. \qquad (~~15~~13\text{-}26)$$

The estimator is obtained by minimizing

$$q = \bar{\mathbf{m}}'\left(\hat{\beta}_{GMM}\right)\mathbf{W}\bar{\mathbf{m}}\left(\hat{\beta}_{GMM}\right),$$

13-37

where $\mathbf{W}$ is a positive definite weighting matrix. The optimal weighting matrix would be

$$\mathbf{W} = \{\text{Asy. Var}[\sqrt{n}\,\bar{\mathbf{m}}(\boldsymbol{\beta})]\}^{-1},$$

which is the inverse of

$$\text{Asy. Var}[\sqrt{n}\bar{\mathbf{m}}(\boldsymbol{\beta})] = \text{Asy. Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i\right] = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma^2\omega_i\mathbf{x}_i\mathbf{x}_i' = \sigma^2\mathbf{Q}^*.$$

[See Section ~~8~~ 9.4.1.] The optimal weighting matrix would be $[\sigma^2\mathbf{Q}^*]^{-1}$. But recall that this minimization problem is an exactly identified case, so the weighting matrix is irrelevant to the solution. You can see the result in the moment equation—that equation is simply the normal equations for ordinary least squares. We can solve the moment equations exactly, so there is no need for the weighting matrix. *Regardless of the covariance matrix of the moments, the GMM estimator for the heteroscedastic regression model is ordinary least squares.* We can use the results we have already obtained to find its asymptotic covariance matrix. The implied estimator is the White estimator in (~~8~~-27). 9 [Once again, 13 see Theorem ~~8~~.2.] The conclusion to be drawn at this point is that until we make some specific assumptions about the variances, we do not have a more efficient estimator than least squares, but we do have to modify the estimated asymptotic covariance matrix.

~~15~~ 13.6.2 SINGLE-EQUATION NONLINEAR MODELS

Suppose that the theory specifies a relationship

$$y_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i,$$

where $\boldsymbol{\beta}$ is a $K \times 1$ parameter vector that we wish to estimate. This may not be a regression relationship, because it is possible that

$$\text{Cov}[\varepsilon_i, h(\mathbf{x}_i, \boldsymbol{\beta})] \neq 0,$$

or even

$$\text{Cov}[\varepsilon_j, \mathbf{x}_j] \neq 0 \text{ for all } i \text{ and } j.$$

Consider, for example, a model that contains lagged dependent variables and autocorrelated disturbances. (See Section ~~19~~ 20.9.3.) For the present, we assume that

$$E[\boldsymbol{\varepsilon}\,|\,\mathbf{X}] \neq \mathbf{0},$$

and

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\,|\,\mathbf{X}] = \sigma^2\boldsymbol{\Omega} = \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is symmetric and positive definite but otherwise unrestricted. The disturbances may be heteroscedastic and/or autocorrelated. But for the possibility of correlation between regressors and disturbances, this model would be a generalized, possibly nonlinear, regression model. Suppose that at each observation i we observe a vector of L variables, $\mathbf{z}_i$, such that $\mathbf{z}_i$ is uncorrelated with ε_i. You will recognize $\mathbf{z}_i$ as a set of **instrumental variables.** The assumptions thus far have implied a set of orthogonality conditions,

XT

$$E[\mathbf{z}_i\varepsilon_i] = \mathbf{0},$$

13-38

which may be sufficient to identify (if $L = K$) or even overidentify (if $L > K$) the parameters of the model. (See Section 8.3.4.)

For convenience, define

$$\mathbf{e}(\mathbf{X}, \hat{\beta}) = y_i - h(\mathbf{x}_i, \hat{\beta}), \quad i = 1, \dots, n,$$

and

$$\mathbf{Z} = n \times L \text{ matrix whose ith row is } \mathbf{z}_i'.$$

By a straightforward extension of our earlier results, we can produce a GMM estimator of β. The sample moments will be

$$\bar{\mathbf{m}}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i e(\mathbf{x}_i, \beta) = \frac{1}{n}\mathbf{Z}'\mathbf{e}(\mathbf{X}, \beta).$$

The minimum distance estimator will be the $\hat{\beta}$ that minimizes

13

$$q = \bar{\mathbf{m}}_n(\hat{\beta})'\mathbf{W}\bar{\mathbf{m}}_n(\hat{\beta}) = \left(\frac{1}{n}[\mathbf{e}(\mathbf{X}, \hat{\beta})'\mathbf{Z}]\right)\mathbf{W}\left(\frac{1}{n}[\mathbf{Z}'\mathbf{e}(\mathbf{X}, \hat{\beta})]\right) \qquad (15\text{-}27)$$

KT for some choice of $\mathbf{W}$ that we have yet to determine. The criterion given earlier produces the **nonlinear instrumental variable estimator.** If we use $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$, then we have 8 exactly the estimation criterion we used in Section 12.7, where we defined the nonlinear instrumental variables estimator. Apparently (15-27) is more general, because we are 13 not limited to this choice of $\mathbf{W}$. For any given choice of $\mathbf{W}$, as long as there are enough orthogonality conditions to identify the parameters, estimation by minimizing q is, at least in principle, a straightforward problem in nonlinear optimization. The optimal choice of $\mathbf{W}$ for this estimator is

13

$$\begin{aligned}\mathbf{W}_{\text{GMM}} &= \{\text{Asy. Var}[\sqrt{n}\,\bar{\mathbf{m}}_n(\beta)]\}^{-1} \\ &= \left\{\text{Asy. Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\varepsilon_i\right]\right\}^{-1} = \left\{\text{Asy. Var}\left[\frac{1}{\sqrt{n}}\mathbf{Z}'\mathbf{e}(\mathbf{X}, \beta)\right]\right\}^{-1}. \qquad (15\text{-}28)\end{aligned}$$

For our model, this is

$$\mathbf{W} = \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}[\mathbf{z}_i\varepsilon_i, \mathbf{z}_j\varepsilon_j]\right]^{-1} = \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\mathbf{z}_i\mathbf{z}_j'\right]^{-1} = \left[\frac{\mathbf{Z}'\Sigma\mathbf{Z}}{n}\right]^{-1}.$$

13

If we insert this result in (15-27), we obtain the criterion for the GMM estimator:

$$q = \left[\left(\frac{1}{n}\right)\mathbf{e}(\mathbf{X}, \hat{\beta})'\mathbf{Z}\right]\left(\frac{\mathbf{Z}'\Sigma\mathbf{Z}}{n}\right)^{-1}\left[\left(\frac{1}{n}\right)\mathbf{Z}'\mathbf{e}(\mathbf{X}, \hat{\beta})\right].$$

There is a possibly difficult detail to be considered. The GMM estimator involves

$$\frac{1}{n}\mathbf{Z}'\Sigma\mathbf{Z} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{z}_i\mathbf{z}_j'\text{Cov}[\varepsilon_i, \varepsilon_j] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{z}_i\mathbf{z}_j'\text{Cov}[(y_i - h(\mathbf{x}_i, \beta)), (y_j - h(\mathbf{x}_j, \beta))].$$

The conditions under which such a double sum might converge to a positive definite matrix are sketched in Section 8.2.2. Assuming that they do hold, estimation appears to require that an estimate of β be in hand already, even though it is the object of estimation.

9

It may be that a consistent but inefficient estimator of β is available. Suppose for the present that one is. If observations are uncorrelated, then the cross-observation terms may be omitted, and what is required is

$$\frac{1}{n}\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'\mathrm{Var}[(y_i - h(\mathbf{x}_i, \beta))].$$

We can use a counterpart to the White (1980) estimator discussed in Section ~~8~~9.4.4 for this case:

$$\mathbf{S}_0 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'(y_i - h(\mathbf{x}_i, \hat{\beta}))^2. \qquad (\cancel{15}13\text{-}29)$$

If the disturbances are autocorrelated but the process is stationary, then Newey and West's (1987a) estimator is available (assuming that the autocorrelations are sufficiently small at a reasonable lag, p):

$$\mathbf{S} = \left[\mathbf{S}_0 + \frac{1}{n}\sum_{\ell=1}^{p} w(\ell)\sum_{i=\ell+1}^{n} e_i e_{i-\ell}(\mathbf{z}_i\mathbf{z}_{i-\ell}' + \mathbf{z}_{i-\ell}\mathbf{z}_i')\right] = \sum_{\ell=0}^{p} w(\ell)\mathbf{S}_\ell, \qquad (\cancel{15}13\text{-}30)$$

where

$$w(\ell) = 1 - \frac{\ell}{p+1}.$$

The maximum lag length p must be determined in advance. We will require that observations that are far apart in time—that is, for which $|i - \ell|$ is large—must have increasingly smaller covariances for us to establish the convergence results that justify OLS, GLS, and now GMM estimation. The choice of p is a reflection of how far back in time one must go to consider the autocorrelation negligible for purposes of estimating $(1/n)\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z}$. Current practice suggests using the smallest integer greater than or equal to $n^{1/4}$.

Still left open is the question of where the initial consistent estimator should be obtained. One possibility is to obtain an inefficient but consistent GMM estimator by using $\mathbf{W} = \mathbf{I}$ in (~~15~~13-27). That is, use a nonlinear (or linear, if the equation is linear) instrumental variables estimator. This first-step estimator can then be used to construct $\mathbf{W}$, which, in turn, can then be used in the GMM estimator. Another possibility is that β may be consistently estimable by some straightforward procedure other than GMM.

Once the GMM estimator has been computed, its asymptotic covariance matrix and asymptotic distribution can be estimated based on Theorem ~~15~~13.2. Recall that

$$\bar{\mathbf{m}}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\varepsilon_i,$$

which is a sum of $L \times 1$ vectors. The derivative, $\partial\bar{\mathbf{m}}_n(\beta)/\partial\beta'$, is a sum of $L \times K$ matrices, so

$$\bar{\mathbf{G}}(\beta) = \partial\bar{\mathbf{m}}(\beta)/\partial\beta' = \frac{1}{n}\sum_{i=1}^{n}\mathbf{G}_i(\beta) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\left[\frac{\partial\varepsilon_i}{\partial\beta'}\right]. \qquad (\cancel{15}13\text{-}31)$$

13-40

In the model we are considering here,

$$\frac{\partial \varepsilon_i}{\partial \boldsymbol{\beta}'} = \frac{-\partial h(\mathbf{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.$$

The derivatives are the pseudoregressors in the linearized regression model that we examined in Section ~~11~~ 7.2.3. Using the notation defined there,

$$\frac{\partial \varepsilon_i}{\partial \boldsymbol{\beta}} = -\mathbf{x}_i^0,$$

so

$$\overline{\mathbf{G}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{G}_i(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} -\mathbf{z}_i\mathbf{x}_i^{0\prime} = -\frac{1}{n}\mathbf{Z}'\mathbf{X}^0. \qquad (\cancel{15}\ 13\text{-32})$$

With this matrix in hand, the estimated asymptotic covariance matrix for the GMM estimator is

$$\text{Est. Asy. Var}[\hat{\boldsymbol{\beta}}] = \left[\overline{\mathbf{G}}(\hat{\boldsymbol{\beta}})'\left(\frac{1}{n}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}\mathbf{Z}\right)^{-1}\overline{\mathbf{G}}(\hat{\boldsymbol{\beta}})\right]^{-1} = [(\mathbf{X}^{0\prime}\mathbf{Z})(\mathbf{Z}'\hat{\boldsymbol{\Sigma}}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X}^0)]^{-1}. \qquad (\cancel{15}\ 13\text{-33})$$

(The two minus signs, a $1/n^2$, and an n^2, all fall out of the result.)

If the $\boldsymbol{\Sigma}$ that appears in (~~15~~ 13-33) were $\sigma^2\mathbf{I}$, then (~~15~~ 13-33) would be precisely the asymptotic covariance matrix that appears in Theorem ~~12~~ 8.2 for linear models and Theorem ~~12~~ 8.3 for nonlinear models. But there is an interesting distinction between this estimator and the IV estimators discussed earlier. In the earlier cases, when there were more instrumental variables than parameters, we resolved the overidentification by specifically choosing a set of K instruments, the K projections of the columns of $\mathbf{X}$ or $\mathbf{X}^0$ into the column space of $\mathbf{Z}$. Here, in contrast, we do not attempt to resolve the overidentification; we simply use all the instruments and minimize the GMM criterion. Now, you should be able to show that when $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ *and we use this information*, when all is said and done, the same parameter estimates will be obtained. But, if we use a weighting matrix that differs from $\mathbf{W} = (\mathbf{Z}'\mathbf{Z}/n)^{-1}$, then they are not.

~~15~~ 13.6.3 SEEMINGLY UNRELATED REGRESSION MODELS

In Section ~~14~~ 10.5, we considered FGLS estimation of the equation system

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{h}_1(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_1,\\
\mathbf{y}_2 &= \mathbf{h}_2(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_2,\\
&\vdots\\
\mathbf{y}_M &= \mathbf{h}_M(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_M.
\end{aligned}$$

The development there extends backwards to the linear system as well. However, none of the estimators considered are consistent if the pseudoregressors, $\mathbf{x}_{tm}^0$, or the actual regressors, $\mathbf{x}_{tm}$ for the linear model, are correlated with the disturbances, ε_{tm}. Suppose we allow for this correlation both within and across equations. (If it is, in fact, absent, then the GMM estimator developed here will remain consistent.) For simplicity in this

13-41

section, we will denote observations with subscript t and equations with subscripts i and j. Suppose, as well, that there are a set of instrumental variables, $\mathbf{z}_t$, such that

$$E[\mathbf{z}_t \varepsilon_{tm}] = \mathbf{0}, \; t = 1, \ldots, T \text{ and } m = 1, \ldots, M. \tag{15-34}$$

(We could allow a separate set of instrumental variables for each equation, but it would needlessly complicate the presentation.)

Under these assumptions, the nonlinear FGLS and ML estimators given earlier will be inconsistent. But a relatively minor extension of the instrumental variables technique developed for the single-equation case in Section 12.4 can be used instead. The sample analog to (15-34) is

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_t [y_{ti} - h_i(\mathbf{x}_t, \boldsymbol{\beta})] = \mathbf{0}, \quad i = 1, \ldots, M.$$

If we use this result for each equation in the system, one at a time, then we obtain exactly the GMM estimator discussed in Section 15.6.2. But, in addition to the efficiency loss that results from not imposing the cross-equation constraints in $\boldsymbol{\beta}$, we would also neglect the correlation between the disturbances. Let

$$\frac{1}{T}\mathbf{Z}'\boldsymbol{\Omega}_{ij}\mathbf{Z} = E\left[\frac{\mathbf{Z}'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\mathbf{Z}}{T}\right]. \tag{15-35}$$

The GMM criterion for estimation in this setting is

$$\begin{aligned} q &= \sum_{i=1}^{M}\sum_{j=1}^{M} [(\mathbf{y}_i - \mathbf{h}_i(\mathbf{X}, \boldsymbol{\beta}))'\mathbf{Z}/T][\mathbf{Z}'\boldsymbol{\Omega}_{ij}\mathbf{Z}/T]^{ij}[\mathbf{Z}'(\mathbf{y}_j - \mathbf{h}_j(\mathbf{X}, \boldsymbol{\beta}))/T] \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} [\boldsymbol{\varepsilon}_i(\boldsymbol{\beta})'\mathbf{Z}/T][\mathbf{Z}'\boldsymbol{\Omega}_{ij}\mathbf{Z}/T]^{ij}[\mathbf{Z}'\boldsymbol{\varepsilon}_j(\boldsymbol{\beta})/T], \end{aligned} \tag{15-36}$$

where $[\mathbf{Z}'\boldsymbol{\Omega}_{ij}\mathbf{Z}/T]^{ij}$ denotes the ijth block of the inverse of the matrix with the ijth block equal to $\mathbf{Z}'\boldsymbol{\Omega}_{ij}\mathbf{Z}/T$. (This matrix is laid out in full in Section 15.6.4.)

GMM estimation would proceed in several passes. To compute any of the variance parameters, we will require an initial consistent estimator of $\boldsymbol{\beta}$. This step can be ~~done with~~ equation-by-equation nonlinear instrumental variables—see Section 12.7—although if equations have parameters in common, then a choice must be made as to which to use. At the next step, the familiar White or Newey-West technique is used to compute, block by block, the matrix in (15-35). Because it is based on a consistent estimator of $\boldsymbol{\beta}$ (we assume), this matrix need not be recomputed. Now, with this result in hand, an iterative solution to the maximization problem in (15-36) can be sought, for example, using the methods of Appendix E. The first-order conditions are

$$\frac{\partial q}{\partial \boldsymbol{\beta}} = -2\sum_{i=1}^{M}\sum_{j=1}^{M} [\mathbf{X}_i^0(\boldsymbol{\beta})'\mathbf{Z}/T][\mathbf{Z}'\mathbf{W}_{ij}\mathbf{Z}/T]^{ij}[\mathbf{Z}'\boldsymbol{\varepsilon}_j(\boldsymbol{\beta})/T] = \mathbf{0}. \tag{15-37}$$

Note again that the blocks of the inverse matrix in the center are extracted from the larger constructed matrix *after inversion*. [This brief discussion might understate the complexity of the optimization problem in (15-36), but that is inherent in the procedure.] At completion, the asymptotic covariance matrix for the GMM estimator is

13-42

estimated with

$$\mathbf{V}_{\text{GMM}} = \frac{1}{T}\left[\sum_{i=1}^{M}\sum_{j=1}^{M}[\mathbf{X}_i^0(\boldsymbol{\beta})'\mathbf{Z}/T][\mathbf{Z}'\mathbf{W}_{ij}\mathbf{Z}/T]^{ij}[\mathbf{Z}'\mathbf{X}_j^0(\boldsymbol{\beta})/T]\right]^{-1}.$$

~~15~~ 13.6.4 SIMULTANEOUS EQUATIONS MODELS WITH HETEROSCEDASTICITY

The GMM estimator in Section ~~15~~ 13.6.1 is, with a minor change of notation, precisely the set of procedures we used in Section ~~13.8~~ 10.5.4 and 10.5.5 to estimate the equations in a simultaneous equations model. Using a GMM estimator, however, will allow us to generalize the covariance structure for the disturbances. We assume that

$$y_{tj} = \mathbf{z}_{tj}'\boldsymbol{\delta}_j + \varepsilon_{tj}, \quad t = 1, \ldots, T,$$

where $\mathbf{z}_{tj} = [\mathbf{Y}_{tj}, \mathbf{x}_{tj}]$. (We use the capital $\mathbf{Y}_{tj}$ to denote the L_j included endogenous variables. Note, as well, that to maintain consistency with Chapter ~~13~~ 10, the roles of the symbols $\mathbf{x}$ and $\mathbf{z}$ are reversed here; $\mathbf{x}$ is now the vector of exogenous variables.) We have assumed that ε_{tj} in the jth equation is neither heteroscedastic nor autocorrelated. There is no need to impose those assumptions at this point. Autocorrelation in the context of a simultaneous equations model is a substantial complication, however. For the present, we will consider the heteroscedastic case only.

The assumptions of the model provide the orthogonality conditions,

$$E[\mathbf{x}_t\varepsilon_{tj}] = E[\mathbf{x}_t(y_{tj} - \mathbf{z}_{tj}'\boldsymbol{\delta}_j)] = \mathbf{0}.$$

If $\mathbf{x}_t$ is taken to be the full set of exogenous variables in the model, then we obtain the criterion for the GMM estimator for the jth equation,

$$\begin{aligned} q &= \left[\frac{\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_j)'\mathbf{X}}{T}\right]\mathbf{W}_{jj}^{-1}\left[\frac{\mathbf{X}'\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_j)}{T}\right] \\ &= \bar{\mathbf{m}}(\boldsymbol{\delta}_j)'\mathbf{W}_{jj}^{-1}\bar{\mathbf{m}}(\boldsymbol{\delta}_j), \end{aligned}$$

where

$$\bar{\mathbf{m}}(\boldsymbol{\delta}_j) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(y_{tj} - \mathbf{z}_{tj}'\boldsymbol{\delta}_j) \quad \text{and} \quad \mathbf{W}_{jj}^{-1} = \text{the GMM weighting matrix.}$$

Once again, this is precisely the estimator defined in Section ~~15~~ 13.6.1. If the disturbances are assumed to be homoscedastic and nonautocorrelated, then the optimal weighting matrix will be an estimator of the inverse of

$$\begin{aligned} \mathbf{W}_{jj} &= \text{Asy. Var}[\sqrt{T}\,\bar{\mathbf{m}}(\boldsymbol{\delta}_j)] \\ &= \text{plim}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'(y_{tj} - \mathbf{z}_{tj}'\boldsymbol{\delta}_j)^2\right] \\ &= \text{plim}\,\frac{1}{T}\sum_{t=1}^{T}\sigma_{jj}\mathbf{x}_t\mathbf{x}_t' \\ &= \text{plim}\,\sigma_{jj}\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right). \end{aligned}$$

13-43

The constant α_{jj} is irrelevant to the solution. If we use $(\mathbf{X}'\mathbf{X})^{-1}$ as the weighting matrix, then the GMM estimator that minimizes q is the 2SLS estimator.

The extension that we can obtain here is to allow for heteroscedasticity of unknown form. There is no need to rederive the earlier result. If the disturbances are heteroscedastic, then

$$\mathbf{W}_{jj} = \operatorname{plim} \frac{1}{T}\sum_{t=1}^{T} \omega_{jj,t}\mathbf{x}_t\mathbf{x}_t' = \operatorname{plim} \frac{\mathbf{X}'\boldsymbol{\Omega}_{jj}\mathbf{X}}{T}.$$

The weighting matrix can be estimated with White's heteroscedasticity consistent estimator—see (15-24)—if a consistent estimator of $\boldsymbol{\delta}_j$ is in hand with which to compute the residuals. One is, because 2SLS ignoring the heteroscedasticity is consistent, albeit inefficient. The conclusion then is that under these assumptions, there is a way to improve on 2SLS by adding another step. The name 3SLS is reserved for the systems estimator of this sort. When choosing between 2.5-stage least squares and Davidson and MacKinnon's suggested "heteroscedastic 2SLS," or **H2SLS**, we chose to opt for the latter. The estimator is based on the initial two-stage least squares procedure. Thus,

AU: KT "H2SLS" is not in chap list

KT

$$\hat{\boldsymbol{\delta}}_{j,\mathrm{H2SLS}} = [\mathbf{Z}_j'\mathbf{X}(\mathbf{S}_{0,jj})^{-1}\mathbf{X}'\mathbf{Z}_j]^{-1}[\mathbf{Z}_j'\mathbf{X}(\mathbf{S}_{0,jj})^{-1}\mathbf{X}'\mathbf{y}_j],$$

where

$$\mathbf{S}_{0,jj} = \sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'(y_{tj} - \mathbf{z}_{tj}'\hat{\boldsymbol{\delta}}_{j,\mathrm{2SLS}})^2.$$

The asymptotic covariance matrix is estimated with

$$\text{Est. Asy. Var}[\hat{\boldsymbol{\delta}}_{j,\mathrm{H2SLS}}] = [\mathbf{Z}_j'\mathbf{X}(\mathbf{S}_{0,jj})^{-1}\mathbf{X}'\mathbf{Z}_j]^{-1}.$$

Extensions of this estimator were suggested by Cragg (1983) and Cumby, Huizinga, and Obstfeld (1983). ~~The H2SLS estimates for Klein's Model I appear in the third panel of results in Table 13.3. Note that the estimated standard errors for these estimates are considerably smaller than those for 2SLS or LIML.~~

AU: Confirm Table 13.3 is correct x-ref.

The GMM estimator for a system of equations is described in Section 15.6.3. As in the single-equation case, a minor change in notation produces the estimators for a simultaneous equations model. As before, we will consider the case of unknown heteroscedasticity only. The extension to autocorrelation is quite complicated. [See Cumby, Huizinga, and Obstfeld (1983).] The orthogonality conditions defined in (15-34) are

$$E[\mathbf{x}_t\varepsilon_{tj}] = E[\mathbf{x}_t(y_{tj} - \mathbf{z}_{tj}'\boldsymbol{\delta}_j)] = \mathbf{0}.$$

If we consider all the equations jointly, then we obtain the criterion for estimation of all the model's parameters,

$$q = \sum_{j=1}^{M}\sum_{l=1}^{M}\left[\frac{\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_j)'\mathbf{X}}{T}\right][\mathbf{W}]^{jl}\left[\frac{\mathbf{X}'\mathbf{e}(\mathbf{z}_t, \boldsymbol{\delta}_l)}{T}\right]$$

$$= \sum_{j=1}^{M}\sum_{l=1}^{M}\bar{\mathbf{m}}(\boldsymbol{\delta}_j)'[\mathbf{W}]^{jl}\bar{\mathbf{m}}(\boldsymbol{\delta}_l),$$

where

$$\bar{\mathbf{m}}(\boldsymbol{\delta}_j) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(y_{tj} - \mathbf{z}_{tj}'\boldsymbol{\delta}_j),$$

13-44

and

$$[\mathbf{W}]^{jl} = \text{block } jl \text{ of the weighting matrix, } \mathbf{W}^{-1}.$$

As before, we consider the optimal weighting matrix obtained as the asymptotic covariance matrix of the empirical moments, $\bar{\mathbf{m}}(\delta_j)$. These moments are stacked in a single vector $\bar{\mathbf{m}}(\delta)$. Then, the jlth block of Asy. Var$[\sqrt{T}\,\bar{\mathbf{m}}(\delta)]$ is

$$\mathbf{\Phi}_{jl} = \operatorname{plim}\left\{\frac{1}{T}\sum_{t=1}^{T}[\mathbf{x}_t\mathbf{x}_t'(y_{tj} - \mathbf{z}_{tj}'\delta_j)(y_{tl} - \mathbf{z}_{tl}'\delta_l)]\right\} = \operatorname{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\omega_{jl,t}\mathbf{x}_t\mathbf{x}_t'\right).$$

If the disturbances are homoscedastic, then $\mathbf{\Phi}_{jl} = \sigma_{jl}[\operatorname{plim}(\mathbf{X}'\mathbf{X}/T)]$ is produced. Otherwise, we obtain a matrix of the form $\mathbf{\Phi}_{jl} = \operatorname{plim}[\mathbf{X}'\mathbf{\Omega}_{jl}\mathbf{X}/T]$. Collecting terms, then, the criterion function for GMM estimation is

$$q = \begin{bmatrix} [\mathbf{X}'(\mathbf{y}_1 - \mathbf{Z}_1\delta_1)]/T \\ [\mathbf{X}'(\mathbf{y}_2 - \mathbf{Z}_2\delta_2)]/T \\ \vdots \\ [\mathbf{X}'(\mathbf{y}_M - \mathbf{Z}_M\delta_M)]/T \end{bmatrix}' \begin{bmatrix} \mathbf{\Phi}_{11} & \mathbf{\Phi}_{12} & \cdots & \mathbf{\Phi}_{1M} \\ \mathbf{\Phi}_{21} & \mathbf{\Phi}_{22} & \cdots & \mathbf{\Phi}_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{\Phi}_{M1} & \mathbf{\Phi}_{M2} & \cdots & \mathbf{\Phi}_{MM} \end{bmatrix}^{-1} \begin{bmatrix} [\mathbf{X}'(\mathbf{y}_1 - \mathbf{Z}_1\delta_1)]/T \\ [\mathbf{X}'(\mathbf{y}_2 - \mathbf{Z}_2\delta_2)]/T \\ \vdots \\ [\mathbf{X}'(\mathbf{y}_M - \mathbf{Z}_M\delta_M)]/T \end{bmatrix}.$$

For implementation, $\mathbf{\Phi}_{jl}$ can be estimated with

$$\hat{\mathbf{\Phi}}_{jl} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'(y_{tj} - \mathbf{z}_{tj}'\mathbf{d}_j)(y_{tl} - \mathbf{z}_{tl}'\mathbf{d}_l),$$

where $\mathbf{d}_j$ is a consistent estimator of δ_j. The two-stage least squares estimator is a natural choice. For the diagonal blocks, this choice is the White estimator as usual. For the off-diagonal blocks, it is a simple extension. With this result in hand, the first-order conditions for GMM estimation are

$$\frac{\partial\hat{q}}{\partial\delta_j} = -2\sum_{l=1}^{M}\left(\frac{\mathbf{Z}_j'\mathbf{X}}{T}\right)\hat{\mathbf{\Phi}}^{jl}\left[\frac{\mathbf{X}'(\mathbf{y}_l - \mathbf{Z}_l\delta_l)}{T}\right],$$

where $\hat{\mathbf{\Phi}}^{jl}$ is the jlth block in the inverse of the estimate of the center matrix in q.

The solution is

$$\begin{bmatrix} \hat{\delta}_{1,\text{GMM}} \\ \hat{\delta}_{2,\text{GMM}} \\ \vdots \\ \hat{\delta}_{M,\text{GMM}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1'\mathbf{X}\hat{\mathbf{\Phi}}^{11}\mathbf{X}'\mathbf{Z}_1 & \mathbf{Z}_1'\mathbf{X}\hat{\mathbf{\Phi}}^{12}\mathbf{X}'\mathbf{Z}_2 & \cdots & \mathbf{Z}_1'\mathbf{X}\hat{\mathbf{\Phi}}^{1M}\mathbf{X}'\mathbf{Z}_M \\ \mathbf{Z}_2'\mathbf{X}\hat{\mathbf{\Phi}}^{21}\mathbf{X}'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{X}\hat{\mathbf{\Phi}}^{22}\mathbf{X}'\mathbf{Z}_2 & \cdots & \mathbf{Z}_2'\mathbf{X}\hat{\mathbf{\Phi}}^{2M}\mathbf{X}'\mathbf{Z}_M \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{Z}_M'\mathbf{X}\hat{\mathbf{\Phi}}^{M1}\mathbf{X}'\mathbf{Z}_1 & \mathbf{Z}_M'\mathbf{X}\hat{\mathbf{\Phi}}^{M2}\mathbf{X}'\mathbf{Z}_2 & \cdots & \mathbf{Z}_M'\mathbf{X}\hat{\mathbf{\Phi}}^{MM}\mathbf{X}'\mathbf{Z}_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{M}\mathbf{Z}_1'\mathbf{X}\hat{\mathbf{\Phi}}^{1j}\mathbf{y}_j \\ \sum_{j=1}^{M}\mathbf{Z}_2'\mathbf{X}\hat{\mathbf{\Phi}}^{2j}\mathbf{y}_j \\ \vdots \\ \sum_{j=1}^{M}\mathbf{Z}_M'\mathbf{X}\hat{\mathbf{\Phi}}^{Mj}\mathbf{y}_j \end{bmatrix}.$$

The asymptotic covariance matrix for the estimator would be estimated with T times the large inverse matrix in brackets.

13-45

Several of the estimators we have already considered are special cases:

- If $\hat{\boldsymbol{\Phi}}_{jj} = \hat{\sigma}_{jj}(\mathbf{X}'\mathbf{X}/T)$ and $\hat{\boldsymbol{\Phi}}_{jl} = \mathbf{0}$ for $j \neq l$, then $\hat{\boldsymbol{\delta}}_j$ is 2SLS.
- If $\hat{\boldsymbol{\Phi}}_{jl} = \mathbf{0}$ for $j \neq l$, then $\hat{\boldsymbol{\delta}}_j$ is H2SLS, the single-equation GMM estimator.
- If $\hat{\boldsymbol{\Phi}}_{jl} = \hat{\sigma}_{jl}(\mathbf{X}'\mathbf{X}/T)$, then $\hat{\boldsymbol{\delta}}_j$ is 3SLS.

As before, the GMM estimator brings efficiency gains in the presence of heteroscedasticity. If the disturbances are homoscedastic, then it is asymptotically the same as 3SLS, [although in a finite sample, it will differ numerically because $\mathbf{S}_{jl}$ will not be identical to $\hat{\sigma}_{jl}(\mathbf{X}'\mathbf{X})$]. ~~These H3SLS estimates for Klein's Model I appear in Table 13.3 with the other full information estimates. As noted there, the sample is too small to fit all three equations jointly, so they are analyzed in pairs.~~

AU: Confirm x-ref to Table 13.3 is correct

13 ~~15~~.6.5 GMM ESTIMATION OF DYNAMIC PANEL DATA MODELS

Panel data are well suited for examining dynamic effects, as in the first-order model,

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \delta y_{i,t-1} + c_i + \varepsilon_{it} \\
&= \mathbf{w}_{it}'\boldsymbol{\theta} + \alpha_i + \varepsilon_{it},
\end{aligned}
$$

where the set of right-hand-side variables, $\mathbf{w}_{it}$, now includes the lagged dependent variable, $y_{i,t-1}$. Adding dynamics to a model in this fashion creates a major change in the interpretation of the equation. Without the lagged variable, the "independent variables" represent the full set of information that produce observed outcome y_{it}. With the lagged variable, we now have in the equation the entire history of the right-hand-side variables, so that any measured influence is conditioned on this history; in this case, any impact of $\mathbf{x}_{it}$ represents the effect of *new* information. Substantial complications arise in estimation of such a model. In both the fixed and random effects settings, the difficulty is that the lagged dependent variable is correlated with the disturbance, even if it is assumed that ε_{it} is not itself autocorrelated. For the moment, consider the fixed effects model as an ordinary regression with a lagged dependent variable. ~~We considered this case in Section 4.9.6 as a regression with a stochastic regressor~~ that is dependent across observations. In that dynamic regression model, the estimator based on T observations is biased in finite samples, but it is consistent in T. ~~That conclusion was the main result of Section 4.9.6.~~ The finite sample bias is of order $1/T$. The same result applies here, but the difference is that whereas before we obtained our large sample results by allowing T to grow large, in this setting, T is assumed to be small and fixed, and large-sample results are obtained with respect to n growing large, not T. The fixed effects estimator of $\boldsymbol{\theta} = [\boldsymbol{\beta}, \delta]$ can be viewed as an average of n such estimators. Assume for now that $T \geq K + 1$ where K is the number of variables in $\mathbf{x}_{it}$. Then, from (9-13), 11

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left[\sum_{i=1}^{n} \mathbf{W}_i'\mathbf{M}^0\mathbf{W}_i\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{W}_i'\mathbf{M}^0\mathbf{y}_i\right] \\
&= \left[\sum_{i=1}^{n} \mathbf{W}_i'\mathbf{M}^0\mathbf{W}_i\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{W}_i'\mathbf{M}^0\mathbf{W}_i\mathbf{d}_i\right] \\
&= \sum_{i=1}^{n} \mathbf{F}_i\mathbf{d}_i,
\end{aligned}
$$

13-46

where the rows of the $T \times (K+1)$ matrix $\mathbf{W}_i$ are $\mathbf{w}'_{it}$ and $\mathbf{M}^0$ is the $T \times T$ matrix that creates deviations from group means [see (9-14)]. Each group-specific estimator, $\mathbf{d}_i$, is inconsistent, as it is biased in finite samples and its variance does not go to zero as n increases. This matrix weighted average of n inconsistent estimators will also be inconsistent. (This analysis is only heuristic. If $T < K+1$, then the individual coefficient vectors cannot be computed.[13])

FN 13

The problem is more transparent in the random effects model. In the model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \delta y_{i,t-1} + u_i + \varepsilon_{it},$$

AU: "Arellano and Bond" is not in chap. list.

the lagged dependent variable is correlated with the compound disturbance in the model, since the same u_i enters the equation for every observation in group i.

Neither of these results renders the model inestimable, but they do make necessary some technique other than our familiar LSDV or FGLS estimators. The general approach, which has been developed in several stages in the literature,[14] relies on instrumental variables estimators and, most recently [by **Arellano and Bond** (1991) and **Arellano and Bover** (1995)] on a **GMM estimator.** For example, in either the fixed or random effects cases, the heterogeneity can be swept from the model by taking first differences, which produces

FN 14

KT

KT

AU: Term "GMM estimator" already KT. Here also?

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + \delta(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1}).$$

This model is still complicated by correlation between the lagged dependent variable and the disturbance (and by its first-order moving average disturbance). But without the group effects, there is a simple instrumental variables estimator available. Assuming that the time series is long enough, one could use the lagged differences, $(y_{i,t-2} - y_{i,t-3})$, or the lagged levels, $y_{i,t-2}$ and $y_{i,t-3}$, as one or two instrumental variables for $(y_{i,t-1} - y_{i,t-2})$. (The other variables can serve as their own instruments.) This is the Anderson and Hsiao estimator developed for this model in Section 12.8.2. ~~By this construction, then, the~~ treatment of this model is a standard application of the instrumental variables technique that we developed in Section 12.8.2.[15] This illustrates the flavor of an instrumental variable approach to estimation. But, as Arellano et al. and Ahn and Schmidt (1995) have shown, there is still more information in the sample that can be brought to bear on estimation, in the context of a GMM estimator, which we now consider.

11.8.2

11.8

FN 15

We can extend the Hausman and Taylor (HT) formulation of the random effects model in Section 12.8.1 to include the lagged dependent variable;

11.5.5

$$\begin{aligned} y_{it} &= \delta y_{i,t-1} + \mathbf{x}'_{1it}\boldsymbol{\beta}_1 + \mathbf{x}'_{2it}\boldsymbol{\beta}_2 + \mathbf{z}'_{1i}\boldsymbol{\alpha}_1 + \mathbf{z}'_{2i}\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i \\ &= \boldsymbol{\theta}'\mathbf{w}_{it} + \varepsilon_{it} + u_i \\ &= \boldsymbol{\theta}'\mathbf{w}_{it} + \eta_{it}, \end{aligned}$$

[13] Further discussion is given by Nickell (1981), Ridder and Wansbeek (1990), and Kiviet (1995).

[14] The model was first proposed in this form by Balestra and Nerlove (1966). See, for example, Anderson and Hsiao (1981, 1982), Bhargava and Sargan (1983), Arellano (1989), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), and Nerlove (2003).

[15] There is a question as to whether one should use differences or levels as instruments. Arellano (1989) and Kiviet (1995) give evidence that the latter is preferable.

where

$$\mathbf{w}_{it} = [y_{i,t-1}, \mathbf{x}'_{1it}, \mathbf{x}'_{2it}, \mathbf{z}'_{1i}, \mathbf{z}'_{2i}]'$$

is now a $(1+K_1+K_2+L_1+L_2)\times 1$ vector. The terms in the equation are the same as in the Hausman and Taylor model. Instrumental variables estimation of the model without the lagged dependent variable is discussed in Section 12.8.1 on the HT estimator. Moreover, by just including $y_{i,t-1}$ in $\mathbf{x}_{2it}$, we see that the HT approach extends to this setting as well, essentially without modification. Arellano et al. suggest a GMM estimator and show that efficiency gains are available by using a larger set of moment conditions. In the previous treatment, we used a GMM estimator constructed as follows: The set of moment conditions we used to formulate the instrumental variables were

11.5.5

$$E\left[\begin{pmatrix}\mathbf{x}_{1it}\\ \mathbf{x}_{2it}\\ \mathbf{z}_{1i}\\ \bar{\mathbf{x}}_{1i.}\end{pmatrix}(\eta_{it}-\bar{\eta}_i)\right] = E\left[\begin{pmatrix}\mathbf{x}_{1it}\\ \mathbf{x}_{2it}\\ \mathbf{z}_{1i}\\ \bar{\mathbf{x}}_{1i.}\end{pmatrix}(\varepsilon_{it}-\bar{\varepsilon}_i)\right] = \mathbf{0}.$$

Section 11.8

This moment condition is used to produce the instrumental variable estimator. We could ignore the nonscalar variance of η_{it} and use simple instrumental variables at this point. However, by accounting for the random effects formulation and using the counterpart to feasible GLS, we obtain the more efficient estimator in (12-29). As usual, this can be done in two steps. The inefficient estimator is computed to obtain the residuals needed to estimate the variance components. This is Hausman and Taylor's steps 1 and 2. Steps 3 and 4 are the GMM estimator based on these estimated variance components.

Arellano et al. suggest that the preceding does not exploit all the information in the sample. In simple terms, within the T observations in group i, we have not used the fact that

$$E\left[\begin{pmatrix}\mathbf{x}_{1it}\\ \mathbf{x}_{2it}\\ \mathbf{z}_{1i}\\ \bar{\mathbf{x}}_{1i.}\end{pmatrix}(\eta_{is}-\bar{\eta}_i)\right] = \mathbf{0} \text{ for some } s \neq t.$$

Thus, for example, not only are disturbances at time t uncorrelated with these variables at time t, arguably, they are uncorrelated with the same variables at time $t-1, t-2$, possibly $t+1$, and so on. In principle, the number of valid instruments is potentially enormous. Suppose, for example, that the set of instruments listed above is strictly exogenous with respect to η_{it} in every period including current, lagged, and future. Then, there are a total of $[T(K_1+K_2)+L_1+K_1)]$ moment conditions for every observation. On this basis alone. Consider, for example, a panel with two periods. We would have for the two periods,

$$E\left[\begin{pmatrix}\mathbf{x}_{1i1}\\ \mathbf{x}_{2i1}\\ \mathbf{x}_{1i2}\\ \mathbf{x}_{2i2}\\ \mathbf{z}_{1i}\\ \bar{\mathbf{x}}_{1i.}\end{pmatrix}(\eta_{i1}-\bar{\eta}_i)\right] = \mathbf{0} \quad \text{and} \quad E\left[\begin{pmatrix}\mathbf{x}_{1i1}\\ \mathbf{x}_{2i1}\\ \mathbf{x}_{1i2}\\ \mathbf{x}_{2i2}\\ \mathbf{z}_{1i}\\ \bar{\mathbf{x}}_{1i.}\end{pmatrix}(\eta_{i2}-\bar{\eta}_i)\right] = \mathbf{0}. \qquad (\cancel{15}\text{-}38)$$

13

How much useful information is brought to bear on estimation of the parameters is uncertain, as it depends on the correlation of the instruments with the included

13-48

FN 16

exogenous variables in the equation. The farther apart in time these sets of variables become the less information is likely to be present. (The literature on this subject contains reference to "strong" versus "weak" instrumental variables.[16]) To proceed, as noted, we can include the lagged dependent variable in $\mathbf{x}_{2i}$. This set of instrumental variables can be used to construct the estimator, actually whether the lagged variable is present or not. We note, at this point, that on this basis, Hausman and Taylor's estimator did not actually use all the information available in the sample. We now have the elements of the Arellano et al. estimator in hand; what remains is essentially the (unfortunately, fairly involved) algebra, which we now develop.

Let

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{w}'_{i1} \\ \mathbf{w}'_{i2} \\ \vdots \\ \mathbf{w}'_{iT} \end{bmatrix} = \text{the full set of rhs data for group } i, \quad \text{and} \quad \mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}.$$

Note that $\mathbf{W}_i$ is assumed to be, a $T \times (1 + K_1 + K_2 + L_1 + L_2)$ matrix. Because there is a lagged dependent variable in the model, it must be assumed that there are actually $T + 1$ observations available on y_{it}. To avoid a cumbersome, cluttered notation, we will leave this distinction embedded in the notation for the moment. Later, when necessary, we will make it explicit. It will reappear in the formulation of the instrumental variables. A total of T observations will be available for constructing the IV estimators. We now form a matrix of instrumental variables. [Different approaches to this have been considered by Hausman and Taylor (1981), Arellano et al. (1991, 1995, 1999), Ahn and Schmidt (1995) and Amemiya and MaCurdy (1986), among others.] We will form a matrix $\mathbf{V}_i$ consisting of $T_i - 1$ rows constructed the same way for $T_i - 1$ observations and a final row that will be different, as discussed later. [This is to exploit a useful algebraic result discussed by Arellano and Bover (1995).] The matrix will be of the form

13

$$\mathbf{V}_i = \begin{bmatrix} \mathbf{v}'_{i1} & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{v}'_{i2} & \cdots & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{a}'_i \end{bmatrix}. \tag{$\cancel{15}$-39}$$

The instrumental variable sets contained in $\mathbf{v}'_{it}$ which have been suggested might include the following from within the model:

$\mathbf{x}_{it}$ and $\mathbf{x}_{i,t-1}$ (i.e., current and one lag of all the time varying variables),
$\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}$ (i.e., all current, past and future values of all the time varying variables),
$\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}$ (i.e., all current and past values of all the time varying variables).

The time-invariant variables that are uncorrelated with u_i, that is $\mathbf{z}_{1i}$, are appended at the end of the nonzero part of each of the first $T - 1$ rows. It may seem that including $\mathbf{x}_2$ in the instruments would be invalid. However, we will be converting the disturbances to deviations from group means which are free of the latent effects—that is, this set of moment conditions will ultimately be converted to what appears in ($\cancel{15}$-38). While the variables are correlated with u_i by construction, they are not correlated with

13

[16]See West (2001).

13-49

$\varepsilon_{it} - \bar{\varepsilon}_i$. The final row of $\mathbf{V}_i$ is important to the construction. Two possibilities have been suggested:

$$\mathbf{a}_i' = [\mathbf{z}_{1i}' \quad \bar{\mathbf{x}}_{i1}] \text{ (produces the Hausman and Taylor estimator)},$$

$$\mathbf{a}_i' = [\mathbf{z}_{1i}' \quad \mathbf{x}_{1i1}', \mathbf{x}_{1i2}', \ldots, \mathbf{x}_{1iT}] \text{ (produces Amemiya and MaCurdy's estimator)}.$$

Note that the $\mathbf{a}$ variables are exogenous time-invariant variables, $\mathbf{z}_{1i}$ and the exogenous time-varying variables, either condensed into the single group mean or in the raw form, with the full set of T observations.

To construct the estimator, we will require a transformation matrix, $\mathbf{H}$, constructed as follows. Let $\mathbf{M}^{01}$ denote the first $T-1$ rows of $\mathbf{M}^0$, the matrix that creates deviations from group means. Then,

$$\mathbf{H} = \begin{bmatrix} \mathbf{M}^{01} \\ \frac{1}{T}\mathbf{i}_T' \end{bmatrix}.$$

Thus, $\mathbf{H}$ replaces the last row of $\mathbf{M}^0$ with a row of $1/T$. The effect is as follows: if $\mathbf{q}$ is T observations on a variable, then $\mathbf{Hq}$ produces $\mathbf{q}^*$ in which the first $T-1$ observations are converted to deviations from group means and the last observation is the group mean. In particular, let the $T \times 1$ column vector of disturbances

$$\eta_i = [\eta_{i1}, \eta_{i2}, \ldots, \eta_{iT}] = [(\varepsilon_{i1} + u_i), (\varepsilon_{i2} + u_i), \ldots, (\varepsilon_{iT} + u_i)]',$$

then

$$\mathbf{H}\eta = \begin{bmatrix} \eta_{i1} - \bar{\eta}_i \\ \vdots \\ \eta_{i,T-1} - \bar{\eta}_i \\ \bar{\eta}_i \end{bmatrix}.$$

We can now construct the moment conditions. With all this machinery in place, we have the result that appears in (15-40), that is 130

$$E[\mathbf{V}_i'\mathbf{H}\eta_i] = E[\mathbf{g}_i] = \mathbf{0}.$$

It is useful to expand this for a particular case. Suppose $T = 3$ and we use as instruments the current values in period 1, and the current and previous values in period 2 and the Hausman and Taylor form for the invariant variables. Then the preceding is

$$E\left[\begin{pmatrix} \mathbf{x}_{1i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{x}_{2i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{z}_{1i} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{1i1} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2i1} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{1i2} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2i2} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{1i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_{1i} \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{x}}_{1i} \end{pmatrix} \begin{pmatrix} \eta_{i1} - \bar{\eta}_i \\ \eta_{i2} - \bar{\eta}_i \\ \bar{\eta}_i \end{pmatrix}\right] = \mathbf{0}. \qquad 13 \ (\cancel{15}\text{-}40)$$

13-50

FN 17

This is the same as (~~15~~13-38).[17] The empirical moment condition that follows from this is

$$\text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}\eta_i$$

$$= \text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}\begin{pmatrix} y_{i1} - \delta y_{i0} - \mathbf{x}_{1i1}'\beta_1 - \mathbf{x}_{2i1}'\beta_2 - \mathbf{z}_{1i}'\alpha_1 - \mathbf{z}_{2i}'\alpha_2 \\ y_{i2} - \delta y_{i1} - \mathbf{x}_{1i2}'\beta_1 - \mathbf{x}_{2i2}'\beta_2 - \mathbf{z}_{1i}'\alpha_1 - \mathbf{z}_{2i}'\alpha_2 \\ \vdots \\ y_{iT} - \delta y_{i,T-1} - \mathbf{x}_{1iT}'\beta_1 - \mathbf{x}_{2iT}'\beta_2 - \mathbf{z}_{1i}'\alpha_1 - \mathbf{z}_{2i}'\alpha_2 \end{pmatrix} = \mathbf{0}.$$

Write this as

$$\text{plim}\frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i = \text{plim}\,\overline{\mathbf{m}} = \mathbf{0}.$$

The GMM estimator $\hat{\theta}$ is then obtained by minimizing

$$q = \overline{\mathbf{m}}'\mathbf{A}\overline{\mathbf{m}}$$

with an appropriate choice of the weighting matrix, $\mathbf{A}$. The optimal weighting matrix will be the inverse of the asymptotic covariance matrix of $\sqrt{n}\,\overline{\mathbf{m}}$. With a consistent estimator of θ in hand, this can be estimated empirically using

$$\text{Est. Asy. Var}[\sqrt{n}\,\overline{\mathbf{m}}] = \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbf{m}}_i\hat{\mathbf{m}}_i' = \frac{1}{n}\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}\hat{\eta}_i\hat{\eta}_i'\mathbf{H}'\mathbf{V}_i.$$

This is a robust estimator that allows an unrestricted $T \times T$ covariance matrix for the T disturbances, $\varepsilon_{it} + u_i$. But, we have assumed that this covariance matrix is the $\boldsymbol{\Sigma}$ defined in (9-28) for the random effects model. To use this information we would, instead, use the residuals in

$$\hat{\eta}_i = \mathbf{y}_i - \mathbf{W}_i\hat{\theta}$$

to estimate σ_u^2 and σ_ε^2 and then $\boldsymbol{\Sigma}$, which produces

$$\text{Est. Asy. Var}[\sqrt{n}\,\overline{\mathbf{m}}] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}\hat{\boldsymbol{\Sigma}}\mathbf{H}'\mathbf{V}_i.$$

We now have the full set of results needed to compute the GMM estimator. The solution to the optimization problem of minimizing q with respect to the parameter vector θ is

$$\hat{\theta}_{GMM} = \left[\left(\sum_{i=1}^{n}\mathbf{W}_i'\mathbf{H}\mathbf{V}_i\right)\left(\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}'\hat{\boldsymbol{\Sigma}}\mathbf{H}\mathbf{V}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}'\mathbf{W}_i\right)\right]^{-1}$$
$$\times\left(\sum_{i=1}^{n}\mathbf{W}_i'\mathbf{H}\mathbf{V}_i\right)\left(\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}'\hat{\boldsymbol{\Sigma}}\mathbf{H}\mathbf{V}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{V}_i'\mathbf{H}'\mathbf{y}_i\right). \qquad \textbf{(~~15~~13-41)}$$

The estimator of the asymptotic covariance matrix for $\hat{\theta}_{GMM}$ is the inverse matrix in brackets.

---

[17] In some treatments [e.g., Blundell and Bond (1998)], an additional condition is assumed for the initial value, y_{i0}, namely $E[y_{i0} \mid \text{exogenous data}] = \mu_0$. This would add a row at the top of the matrix in (~~15~~13-40) containing $[(y_{i0} - \mu_0), 0, 0]$.

13-51

The remaining loose end is how to obtain the consistent estimator of $\hat{\theta}$ to compute $\hat{\Sigma}$. Recall that the GMM estimator is consistent with any positive definite weighting matrix, $\mathbf{A}$, in our preceding expression. Therefore, for an initial estimator, we could set $\mathbf{A} = \mathbf{I}$ and use the simple instrumental variables estimator,

$$\hat{\theta}_{\text{IV}} = \left[\left(\sum_{i=1}^{n} \mathbf{W}_i'\mathbf{H}\mathbf{V}_i\right)\left(\sum_{i=1}^{n} \mathbf{V}_i'\mathbf{H}\mathbf{W}_i\right)\right]^{-1}\left(\sum_{i=1}^{n} \mathbf{W}_i'\mathbf{H}\mathbf{V}_i\right)\left(\sum_{i=1}^{n} \mathbf{V}_i'\mathbf{H}\mathbf{y}_i\right).$$

It is more common to proceed directly to the "two-stage least squares" estimator (see Sections 12.3.3 and 12.8.2), which uses

8.3.4 11.12.3

$$\mathbf{A} = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{V}_i'\mathbf{H}'\mathbf{H}\mathbf{V}_i\right)^{-1}.$$

13

The estimator is, then, the one given earlier in (~~15~~-41) with $\hat{\Sigma}$ replaced by $\mathbf{I}_T$. Either estimator is a function of the sample data only and provides the initial estimator we need.

Ahn and Schmidt (among others) observed that the IV estimator proposed here, as extensive as it is, still neglects quite a lot of information and is therefore (relatively) inefficient. For example, in the first differenced model,

$$E[y_{is}(\varepsilon_{it} - \varepsilon_{i,t-1})] = 0, \quad s = 0, \ldots, t-2, \quad t = 2, \ldots, T.$$

FN 18

That is, the *level* of y_{is} is uncorrelated with the differences of disturbances that are at least two periods subsequent.[18] (The differencing transformation, as the transformation to deviations from group means, removes the individual effect.) The corresponding moment equations that can enter the construction of a GMM estimator are

$$\frac{1}{n}\sum_{i=1}^{n} y_{is}[(y_{it} - y_{i,t-1}) - \delta(y_{i,t-1} - y_{i,t-2}) - (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta}] = 0$$
$$s = 0, \ldots, t-2, \quad t = 2, \ldots, T.$$

Altogether, Ahn and Schmidt identify $T(T-1)/2 + T - 2$ such equations that involve mixtures of the levels and differences of the variables. The main conclusion that they demonstrate is that in the dynamic model, there is a large amount of information to be gleaned not only from the familiar relationships among the levels of the variables but also from the implied relationships between the levels and the first differences. The issue of correlation between the transformed y_{it} and the deviations of ε_{it} is discussed in the papers cited. [As Ahn and Schmidt show, there are potentially huge numbers of additional orthogonality conditions in this model owing to the relationship between first differences and second moments. We do not consider those. The matrix $\mathbf{V}_i$ could be huge. Consider a model with 10 time-varying right-hand-side variables and suppose T_i is 15. Then, there are 15 rows and roughly 15 × (10 × 15) or 2,250 columns. The Ahn and Schmidt estimator, which involves potentially thousands of instruments in a model containing only a handful of parameters, may become a bit impractical at this point. The

[18] This is the approach suggested by Holtz-Eakin (1988) and Holtz-Eakin, Newey, and Rosen (1988).

[13-52]

common approach is to use only a small subset of the available instrumental variables. The order of the computation grows as the number of parameters times the square of T.]

The number of orthogonality conditions (instrumental variables) used to estimate the parameters of the model is determined by the number of variables in $\mathbf{v}_{it}$ and $\mathbf{a}_i$ in (15-39). In most cases, the model is vastly overidentified—there are far more orthogonality conditions than parameters. As usual in GMM estimation, a test of the overidentifying restrictions can be based on q, the estimation criterion. At its minimum, the limiting distribution of nq is chi-squared with degrees of freedom equal to the number of instrumental variables in total minus $(1 + K_1 + K_2 + L_1 + L_2)$.[19]

[13C] [FN 19]

Example ~~15~~13.10 GMM Estimation of a Dynamic Panel Data Model of Local Government Expenditures

Dahlberg and Johansson (2000) estimated a model for the local government expenditure of several hundred municipalities in Sweden observed over the nine-year period $t = 1979$ to 1987. The equation of interest is

$$S_{i,t} = \alpha_t + \sum_{j=1}^{m} \beta_j S_{i,t-j} + \sum_{j=1}^{m} \gamma_j R_{i,t-j} + \sum_{j=1}^{m} \delta_j G_{i,t-j} + f_i + \varepsilon_{it},$$

for $i = 1, \ldots, n = 265$, and $t = m + 1, \ldots, 9$. (We have changed their notation slightly to make it more convenient.) $S_{i,t}$, $R_{i,t}$, and $G_{i,t}$ are municipal spending, receipts (taxes and fees), and central government grants, respectively. Analogous equations are specified for the current values of $R_{i,t}$ and $G_{i,t}$. The appropriate lag length, m, is one of the features of interest to be determined by the empirical study. The model contains a municipality specific effect, f_i, which is not specified as being either "fixed" or "random." To eliminate the individual effect, the model is converted to first differences. The resulting equation is

$$\Delta S_{i,t} = \lambda_t + \sum_{j=1}^{m} \beta_j \Delta S_{i,t-j} + \sum_{j=1}^{m} \gamma_j \Delta R_{i,t-j} + \sum_{j=1}^{m} \delta_j \Delta G_{i,t-j} + u_{it},$$

or

$$y_{i,t} = \mathbf{x}_{i,t}'\theta + u_{i,t},$$

where $\Delta S_{i,t} = S_{i,t} - S_{i,t-1}$ and so on and $u_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$. This removes the group effect and leaves the time effect. Because the time effect was unrestricted to begin with, $\Delta\alpha_t = \lambda_t$ remains an unrestricted time effect, which is treated as "fixed" and modeled with a time-specific dummy variable. The maximum lag length is set at $m = 3$. With nine years of data, this leaves usable observations from 1983 to 1987 for estimation, that is, $t = m + 2, \ldots, 9$. Similar equations were fit for $R_{i,t}$ and $G_{i,t}$.

The orthogonality conditions claimed by the authors are

$$E[S_{i,s}u_{i,t}] = E[R_{i,s}u_{i,t}] = E[G_{i,s}u_{i,t}] = 0, \quad s = 1, \ldots, t-2.$$

The orthogonality conditions are stated in terms of the levels of the financial variables and the differences of the disturbances. The issue of this formulation as opposed to, for example, $E[\Delta S_{i,s}\Delta\varepsilon_{i,t}] = 0$ (which is implied) is discussed by Ahn and Schmidt (1995). As we shall see, this set of orthogonality conditions implies a total of 80 instrumental variables. The authors use only the first of the three sets listed, which produces a total of 30. For the five observations, using the formulation developed in Section 15.6.5, we have the following matrix

[AQ: Is x-ref to Section 15.6.5 OK?]

[19] This is true generally in GMM estimation. It was proposed for the dynamic panel data model by Bhargava and Sargan (1983).

13-53

of instrumental variables for the orthogonality conditions

$$\mathbf{Z}_i = \begin{array}{c} \left[\begin{array}{cccccccccc} S_{81-79} & d_{83} & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 \\ \mathbf{0}' & 0 & S_{82-79} & d_{84} & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 \\ \mathbf{0}' & 0 & \mathbf{0}' & 0 & S_{83-79} & d_{85} & \mathbf{0}' & 0 & \mathbf{0}' & 0 \\ \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & S_{84-79} & d_{86} & \mathbf{0}' & 0 \\ \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & S_{85-79} & d_{87} \end{array}\right] \end{array} \begin{array}{l} 1983 \\ 1984 \\ 1985 \\ 1986 \\ 1987 \end{array}$$

where the notation S_{t1-t0} indicates the range of years for that variable. For example, S_{83-79} denotes $[S_{i,1983}, S_{i,1982}, S_{i,1981}, S_{i,1980}, S_{i,1979}]$ and d_{year} denotes the year-specific dummy variable. Counting columns in $\mathbf{Z}_i$ we see that using only the lagged values of the dependent variable and the time dummy variables, we have $(3+1)+(4+1)+(5+1)+(6+1)+(7+1)=30$ instrumental variables. Using the lagged values of the other two variables in each equation would add 50 more, for a total of 80 if all the orthogonality conditions suggested earlier were employed. Given the preceding construction, the orthogonality conditions are now

$$E[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0},$$

where $\mathbf{u}_i = [u_{i,1987}, u_{i,1986}, u_{i,1985}, u_{i,1984}, u_{i,1983}]'$. The empirical moment equation is

1983 1984 1986 1987

$$\text{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{u}_i\right] = \text{plim}\,\overline{\mathbf{m}}(\theta) = \mathbf{0}.$$

The parameters are vastly overidentified. Using only the lagged values of the dependent variable in each of the three equations estimated, there are 30 moment conditions and 14 parameters being estimated when $m=3$, 11 when $m=2$, 8 when $m=1$, and 5 when $m=0$. (As we do our estimation of each of these, we will retain the same matrix of instrumental variables in each case.) GMM estimation proceeds in two steps. In the first step, basic, unweighted instrumental variables is computed using

$$\hat{\theta}'_{IV} = \left[\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{Z}_i\right)\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1}\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{Z}_i\right)\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{y}_i\right),$$

where

$$\mathbf{y}_i' = (\Delta S_{83} \quad \Delta S_{84} \quad \Delta S_{85} \quad \Delta S_{86} \quad \Delta S_{87}),$$

and

$$\mathbf{X}_i = \begin{bmatrix} \Delta S_{82} & \Delta S_{81} & \Delta S_{80} & \Delta R_{82} & \Delta R_{81} & \Delta R_{80} & \Delta G_{82} & \Delta G_{81} & \Delta G_{80} & 1 & 0 & 0 & 0 & 0 \\ \Delta S_{83} & \Delta S_{82} & \Delta S_{81} & \Delta R_{83} & \Delta R_{82} & \Delta R_{81} & \Delta G_{83} & \Delta G_{82} & \Delta G_{81} & 0 & 1 & 0 & 0 & 0 \\ \Delta S_{84} & \Delta S_{83} & \Delta S_{82} & \Delta R_{84} & \Delta R_{83} & \Delta R_{82} & \Delta G_{84} & \Delta G_{83} & \Delta G_{82} & 0 & 0 & 1 & 0 & 0 \\ \Delta S_{85} & \Delta S_{84} & \Delta S_{83} & \Delta R_{85} & \Delta R_{84} & \Delta R_{83} & \Delta G_{85} & \Delta G_{84} & \Delta G_{83} & 0 & 0 & 0 & 1 & 0 \\ \Delta S_{86} & \Delta S_{85} & \Delta S_{84} & \Delta R_{86} & \Delta R_{85} & \Delta R_{84} & \Delta G_{86} & \Delta G_{85} & \Delta G_{84} & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The second step begins with the computation of the new weighting matrix,

$$\hat{\mathbf{\Phi}} = \text{Est. Asy. Var}[\sqrt{n}\,\overline{\mathbf{m}}] = \frac{1}{N}\sum_{i=1}^{n}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i.$$

13-54

After multiplying and dividing by the implicit $(1/n)$ in the outside matrices, we obtain the estimator,

$$\begin{aligned}\theta'_{GMM} &= \left[\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{Z}_i\right)\left(\sum_{i=1}^{n}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1}\\ &\quad\times\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{Z}_i\right)\left(\sum_{i=1}^{n}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{y}_i\right)\\ &= \left[\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1}\left(\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}\left(\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{y}_i\right).\end{aligned}$$

The estimator of the asymptotic covariance matrix for the estimator is the inverse matrix in square brackets in the first line of the result.

The primary focus of interest in the study was not the estimator itself, but the lag length and whether certain lagged values of the independent variables appeared in each equation. These restrictions would be tested by using the GMM criterion function, which in this formulation would be (based on recomputing the residuals after GMM estimation)

$$n\,q = \left(\sum_{i=1}^{n}\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)\mathbf{W}\left(\sum_{i=1}^{n}\mathbf{Z}_i'\hat{\mathbf{u}}_i\right).$$

Note that the weighting matrix is not (necessarily) recomputed. For purposes of testing hypotheses, the same weighting matrix should be used.

At this point, we will consider the appropriate lag length, m. The specification can be reduced simply by redefining $\mathbf{X}$ to change the lag length. To test the specification, the weighting matrix must be kept constant for all restricted versions ($m = 2$ and $m = 1$) of the model.

13 The Dahlberg and Johansson data may be downloaded from the *Journal of Applied Econometrics* Web site—see Appendix Table F~~15~~13.1. The authors provide the summary statistics for the raw data that are given in Table ~~15~~13.3. The data used in the study and provided in the internet source are nominal values in Swedish Kroner, deflated by a municipality-specific price index then converted to per capita values. Descriptive statistics for the raw data appear in Table ~~15~~13.3.[20] Equations were estimated for all three variables, with maximum lag lengths of $m = 1$, 2, and 3. (The authors did not provide the actual estimates.) Estimation is done using the methods developed by Ahn and Schmidt (1995), Arellano and Bover (1995), and Holtz-Eakin, Newey, and Rosen (1988), as described. The estimates of the first specification provided are given in Table ~~15~~13.4.

Table ~~15~~13.5 contains estimates of the model parameters for each of the three equations, and for the three lag lengths, as well as the value of the GMM criterion function for each model estimated. The base case for each model has $m = 3$. There are three restrictions implied by

TABLE ~~15~~13.3 Descriptive Statistics for Local Expenditure Data

Variable	Mean	Std. Deviation	Minimum	Maximum
Spending	18478.51	3174.36	12225.68	33883.25
Revenues	13422.56	3004.16	6228.54	29141.62
Grants	5236.03	1260.97	1570.64	12589.14

[20] The data provided on the Web site and used in our computations were further transformed by dividing by 100,000.

13-55

TABLE ~~15~~ 13.4 Estimated Spending Equation

Variable	*Estimate*	*Standard Error*	*t Ratio*
Year 1983	−0.0036578	0.0002969	−12.32
Year 1984	−0.00049670	0.0004128	−1.20
Year 1985	0.00038085	0.0003094	1.23
Year 1986	0.00031469	0.0003282	0.96
Year 1987	0.00086878	0.0001480	5.87
Spending $(t-1)$	1.15493	0.34409	3.36
Revenues $(t-1)$	−1.23801	0.36171	−3.42
Grants $(t-1)$	0.016310	0.82419	0.02
Spending $(t-2)$	−0.0376625	0.22676	−0.17
Revenues $(t-2)$	0.0770075	0.27179	0.28
Grants $(t-2)$	1.55379	0.75841	2.05
Spending $(t-3)$	−0.56441	0.21796	−2.59
Revenues $(t-3)$	0.64978	0.26930	2.41
Grants $(t-3)$	1.78918	0.69297	2.58

each reduction in the lag length. The critical chi-squared value for three degrees of freedom is 7.81 for 95 percent significance, so at this level, we find that the two-level model is just barely accepted for the spending equation, but clearly appropriate for the other two—the difference between the two criteria is 7.62. Conditioned on $m = 2$, only the revenue model rejects the restriction of $m = 1$. As a final test, we might ask whether the data suggest that perhaps no lag structure at all is necessary. The GMM criterion value for the three equations with only the time dummy variables are 45.840, 57.908, and 62.042, respectively. Therefore, all three zero lag models are rejected.

Among the interests in this study were the appropriate critical values to use for the specification test of the moment restriction. With 16 degrees of freedom, the critical chi-squared value for 95 percent significance is 26.3, which would suggest that the revenues equation is misspecified. Using a bootstrap technique, the authors find that a more appropriate critical value leaves the specification intact. Finally, note that the three-equation model in the $m = 3$ columns of Table 15.5 imply a **vector autoregression** of the form

132

$$\mathbf{y}_t = \boldsymbol{\Gamma}_1\mathbf{y}_{t-1} + \boldsymbol{\Gamma}_2\mathbf{y}_{t-2} + \boldsymbol{\Gamma}_3\mathbf{y}_{t-3} + \mathbf{v}_t,$$

where $\mathbf{y}_t = (\Delta S_t, \Delta R_t, \Delta G_t)'$. We will explore the properties and characteristics of equation systems such as this in our discussion of time-series models in Chapter ~~21~~ 22.

TABLE ~~15~~ 13.5 Estimated Lag Equations for Spending, Revenue, and Grants

	Expenditure Model			*Revenue Model*			*Grant Model*		
	m = 3	m = 2	m = 1	m = 3	m = 2	m = 1	m = 3	m = 2	m = 1
S_{t-1}	1.155	0.8742	0.5562	−0.1715	−0.3117	−0.1242	−0.1675	−0.1461	−0.1958
S_{t-2}	−0.0377	0.2493	—	0.1621	−0.0773	—	−0.0303	−0.0304	—
S_{t-3}	−0.5644	—	—	−0.1772	—	—	−0.0955	—	—
R_{t-1}	−1.2380	−0.8745	−0.5328	−0.0176	0.1863	−0.0245	0.1578	0.1453	0.2343
R_{t-2}	0.0770	−0.2776	—	−0.0309	0.1368	—	0.0485	0.0175	—
R_{t-3}	0.6497	—	—	0.0034	—	—	0.0319	—	—
G_{t-1}	0.0163	−0.4203	0.1275	−0.3683	0.5425	−0.0808	−0.2381	−0.2066	−0.0559
G_{t-2}	1.5538	0.1866	—	−2.7152	2.4621	—	−0.0492	−0.0804	—
G_{t-3}	1.7892	—	—	0.0948	—	—	0.0598	—	—
n q	22.8287	30.4526	34.4986	30.5398	34.2590	53.2506	17.5810	20.5416	27.5927

delete minus

13-56

13 ~~15~~.7 SUMMARY AND CONCLUSIONS

The generalized method of moments provides an estimation framework that includes least squares, nonlinear least squares, instrumental variables, and maximum likelihood, and a general class of estimators that extends beyond these. But it is more than just a theoretical umbrella. The GMM provides a method of formulating models and implied estimators without making strong distributional assumptions. Hall's model of household consumption is a useful example that shows how the optimization conditions of an underlying economic theory produce a set of distribution-free estimating equations. In this chapter, we first examined the classical method of moments. GMM as an estimator is an extension of this strategy that allows the analyst to use additional information beyond that necessary to identify the model, in an optimal fashion. After defining and establishing the properties of the estimator, we then turned to inference procedures. It is convenient that the GMM procedure provides counterparts to the familiar trio of test statistics: Wald, LM, and LR. In the final section, we specialized the GMM estimator for linear and nonlinear equations and multiple-equation models. We then developed an example that appears at many points in the recent applied literature, the dynamic panel data model with individual specific effects, and lagged values of the dependent variable.

AU: Terms with blue checks were not bold KTs in chapter

Key Terms and Concepts

- Analog estimation
- Arellano and Bover estimator
- Central limit theorem
- ~~Central moments~~
- Criterion function
- Dynamic panel data model
- Empirical moment equation
- Ergodic theorem
- Euler equation
- Exactly identified cases
- ~~Exactly defined cases~~
- Exponential family
- Generalized method of moments
- GMM estimator
- ~~Identification~~
- Instrumental variables
- Likelihood ratio statistic
- LM statistic
- Martingale difference series
- Maximum likelihood estimator
- Mean value theorem
- Method of moment generating functions
- Method of moments
- Method of moments estimators
- Minimum distance estimator (MDE)
- Moment equation
- Newey-West estimator
- Nonlinear instrumental variable estimator
- Optimal weighting matrix
- Order condition
- Orthogonality conditions
- Overidentifying restrictions
- Overidentified cases
- Population moment equation
- Probability limit
- Random sample
- Rank condition
- Slutsky theorem
- Specification test
- Sufficient statistic
- Taylor series
- Uncentered moment
- ~~Vector autoregression~~
- Wald statistic
- Weighted least squares
- Weighting matrix

Exercises

1. For the normal distribution $\mu_{2k} = \sigma^{2k}(2k)!/(k!2^k)$ and $\mu_{2k+1} = 0$, $k = 0, 1, \ldots$. Use this result to analyze the two estimators

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \quad \text{and} \quad b_2 = \frac{m_4}{m_2^2},$$

13-57
End 13

where $m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$. The following result will be useful:

$$\text{Asy. Cov}[\sqrt{n}m_j, \sqrt{n}m_k] = \mu_{j+k} - \mu_j\mu_k + jk\mu_2\mu_{j-1}\mu_{k-1} - j\mu_{j-1}\mu_{k+1} - k\mu_{k-1}\mu_{j+1}.$$

Use the delta method to obtain the asymptotic variances and covariance of these two functions, assuming the data are drawn from a normal distribution with mean μ and variance σ^2. (Hint: Under the assumptions, the sample mean is a consistent estimator of μ, so for purposes of deriving asymptotic results, the difference between $\bar{x}$ and μ may be ignored. As such, no generality is lost by assuming the mean is zero, and proceeding from there.) Obtain $\mathbf{V}$, the 3 × 3 covariance matrix for the three moments, then use the delta method to show that the covariance matrix for the two estimators is

and

$$\mathbf{JVJ}' = \begin{bmatrix} 6/n & 0 \\ 0 & 24/n \end{bmatrix},$$

where $\mathbf{J}$ is the 2 × 3 matrix of derivatives.

2. Using the results in Example ~~15.7~~ 13, estimate the asymptotic covariance matrix of the method of moments estimators of P and λ based on m_1' and m_2'. [Note: You will need to use the data in Example C.1 to estimate $\mathbf{V}$.]
3. **Exponential Families of Distributions.** For each of the following distributions, determine whether it is an exponential family by examining the log-likelihood function. Then, identify the sufficient statistics.
   a. Normal distribution with mean μ and variance σ^2.
   b. The Weibull distribution in Exercise 4 in Chapter ~~16~~ 14.
   c. The mixture distribution in Exercise 3 in Chapter ~~16~~ 14.
4. In the classical regression model with heteroscedasticity, which is more efficient, ordinary least squares or GMM? Obtain the two estimators and their respective asymptotic covariance matrices, then prove your assertion.
5. Consider the probit model analyzed in Chapter ~~23~~ 17. The model states that for given vector of independent variables,

$$\text{Prob}[y_i = 1 \mid \mathbf{x}_i] = \Phi[\mathbf{x}_i'\boldsymbol{\beta}], \quad \text{Prob}[y_i = 0 \mid \mathbf{x}_i] = 1 - \text{Prob}[y_i = 1 \mid \mathbf{x}_i].$$

Consider a GMM estimator based on the result that

$$E[y_i \mid \mathbf{x}_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta}).$$

This suggests that we might base estimation on the orthogonality conditions

$$E[(y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i] = \mathbf{0}.$$

Construct a GMM estimator based on these results. Note that this is not the nonlinear least squares estimator. Explain—what would the orthogonality conditions be for nonlinear least squares estimation of this model?

6. Consider GMM estimation of a regression model as shown at the beginning of Example ~~15.8~~ 13. Let $\mathbf{W}_1$ be the optimal weighting matrix based on the moment equations. Let $\mathbf{W}_2$ be some other positive definite matrix. Compare the asymptotic covariance matrices of the two proposed estimators. Show conclusively that the asymptotic covariance matrix of the estimator based on $\mathbf{W}_1$ is not larger than that based on $\mathbf{W}_2$.

Au: Is exercise title OK set bold (rather than italics)?