

Applications of Generalized Method of Moments Estimation

Jeffrey M. Wooldridge

The method of moments approach to parameter estimation dates back more than 100 years (Stigler, 1986). The notion of a moment is fundamental for describing features of a population. For example, the population mean (or population average), usually denoted μ , is the moment that measures central tendency. If y is a random variable describing the population of interest, we also write the population mean as $E(y)$, the expected value or mean of y . (The mean of y is also called the first moment of y .) The population variance, usually denoted σ^2 or $\text{Var}(y)$, is defined as the second moment of y centered about its mean: $\text{Var}(y) = E[(y - \mu)^2]$. The variance, also called the second central moment, is widely used as a measure of spread in a distribution.

Since we can rarely obtain information on an entire population, we use a sample from the population to estimate population moments. If $\{y_i; i = 1, \dots, n\}$ is a sample from a population with mean μ , the *method of moments estimator* of μ is just the sample average: $\bar{y} = (y_1 + y_2 + \dots + y_n)/n$. Under random sampling, \bar{y} is unbiased and consistent for μ regardless of other features of the underlying population. Further, as long as the population variance is finite, \bar{y} is the best linear unbiased estimator of μ . An unbiased and consistent estimator of σ^2 also exists and is called the sample variance, usually denoted s^2 .¹

Method of moments estimation applies in more complicated situations. For example, suppose that in a population with $\mu > 0$, we know that the variance is three times the mean: $\sigma^2 = 3\mu$. The sample average, \bar{y} , is unbiased and consistent

¹ See Wooldridge (2000, appendix C) for more discussion of the sample mean and sample variance as method of moments estimators.

■ *Jeffrey M. Wooldridge is University Distinguished Professor of Economics, Michigan State University, East Lansing, Michigan. His e-mail address is (wooldri1@msu.edu).*

for μ , but so is a different estimator, namely, $s^2/3$. The existence of two unbiased, consistent method of moments estimators raises an obvious question: Which should we use? One possible answer is to choose the estimator with the smallest sampling variance, so that we obtain the most precise estimator of μ . But in this case, it turns out that for some values of μ , the sample average has a smaller variance, while for other values, $s^2/3$ has the smaller variance. Is there an estimator that combines the information in \bar{y} and $s^2/3$ and performs better than either would alone? Yes, provided we restrict ourselves to large-sample comparisons. The theory of *generalized method of moments* (GMM) tells us how to use the two sets of *population moment conditions*, which in this case can be written as $E(y) = \mu$ and $E[(y - \mu)^2] = 3\mu$, in a manner that minimizes the asymptotic variance among method of moments estimators of μ .

The preceding setup illustrates two features that are common in applications of generalized method of moments. First, we have two population moment conditions but only one parameter, μ , to estimate. If we replace the population moments $E(y)$ and $E[(y - \mu)^2]$ with their sample counterparts, we obtain two equations in one unknown, the estimate $\hat{\mu}$. The two sample equations can be written as $\bar{y} = \hat{\mu}$ and $[(y_1 - \hat{\mu})^2 + \cdots + (y_n - \hat{\mu})^2]/n = 3\hat{\mu}$. Generally, there is no value of $\hat{\mu}$ that solves both of these equations. Instead, GMM weights the two sample moment conditions to obtain an asymptotically optimal estimator. A second noteworthy feature of this example is that at least one moment condition is nonlinear in the parameter, μ , something that is common in advanced applications of GMM.²

In estimating the parameters of a population regression function, a parallel situation can arise. When the error term is heteroskedastic, it is generally possible to add moment conditions to those used by ordinary least squares and obtain an asymptotically more efficient estimator. The key is to weight the entire set of moment conditions in an optimal way.

Method of Moments Estimators: From Ordinary Least Squares to Generalized Method of Moments

Many commonly used estimators in econometrics, including ordinary least squares and instrumental variables, are derived most naturally using the method of moments. As a starting point, consider a population linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u,$$

² Some authors prefer not to make a distinction between method of moments and “generalized” method of moments. Early applications of the method of moments were to estimate the parameters of univariate population distributions where the number of moment conditions was equal to the number of parameters to be estimated. In these applications, the moment equations usually could be solved in closed form. In addition to allowing more moments than parameters to estimate, GMM was constructed directly for econometric applications with complicated models and data structures.

where y is the dependent or response variable, the x_j are the covariates or explanatory variables, and u is the unobserved error or disturbance. The goal is to estimate the $k + 1$ regression parameters, β_j , given a random sample on $(y, x_1, x_2, \dots, x_k)$.

A common assumption in linear regression is that the population error has a mean of zero and that each x_j is uncorrelated with the error term, that is,

$$E(u) = 0, \quad E(x_j u) = 0, \quad j = 1, \dots, k.$$

For brevity, we call this the “zero correlation assumption.” This assumption implies that $k + 1$ population moments involving the covariates and the error are identically zero. If we write the error in terms of the observable variables and unknown parameters as $u = y - \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_k x_k$, and we replace the population moments with their sample counterparts, the moment conditions implied by the zero correlation assumption lead to the first-order conditions for the ordinary least squares estimator; see, for example, Wooldridge (2000, equation 3.13).

The zero correlation assumption is the weakest sense in which the covariates are exogenous in the population linear model. If these assumptions are the only ones we are willing to make, ordinary least squares is the only sensible estimator of the β_j .³

Often we are willing to make a stronger exogeneity assumption. If we assume that the error term has a zero mean conditional on the covariates,

$$E(u|x_1, x_2, \dots, x_k) = 0,$$

alternatives to ordinary least squares become available. Why? Because the zero conditional mean assumption ensures that *any* function of the covariates is uncorrelated with u . For example, all functions of the form $x_h x_j$, $h, j = 1, \dots, k$, are uncorrelated with u , even though these squares and interactions do not appear in the original model. If the model is a wage equation and the covariates include education and experience, the zero conditional mean assumption implies that the error is uncorrelated with the squares of education and experience and an interaction between them, even if these functions are not in the original model.

Under the zero conditional mean assumption, is it possible, by adding zero correlation assumptions involving nonlinear functions of the x_j , to improve upon the ordinary least squares estimator? The answer is yes, provided there is heteroskedasticity. Specifically, if the zero conditional mean assumption holds and $\text{Var}(u|x_1, \dots, x_k)$ depends on some of the covariates, it is possible to obtain

³ Other common estimators, such as weighted least squares or the maximum likelihood estimator under an assumed distribution for u given x , are generally inconsistent if only the zero correlation assumption holds.

method of moments estimators that have smaller asymptotic variances than the ordinary least squares estimator.⁴

Cragg (1983) was the first to discover that one can improve over ordinary least squares in the presence of heteroskedasticity of unknown form by applying generalized method of moments. How does GMM work in this case? First, one must decide which extra moment conditions to add to those generated by the usual zero correlation assumption. Next, having first done ordinary least squares, one must obtain the *weighting matrix* that is a crucial component to an efficient GMM analysis. The weighting matrix is obtained by inverting a consistent estimator of the variance-covariance matrix of the moment conditions. If there are $m > k + 1$ total moment conditions, where k is the number of covariates in the model, then the weighting matrix has dimension $m \times m$. The GMM estimator minimizes a quadratic form in the sample moment conditions, where the weighting matrix appears in the quadratic form. As shown by Hansen (1982) and White (1982), this choice of the weighting matrix is asymptotically optimal.⁵ The intuition behind the optimality of this weighting matrix is easiest when the moment conditions are uncorrelated with one another.⁶ Then, the weighting matrix can be taken to be a diagonal matrix, where each diagonal element is the reciprocal of the variance of the corresponding moment condition. In other words, moment conditions with larger variances receive relatively less weight in the estimation, since they contain less information about the population parameters. Moment conditions with smaller variances receive relatively more weight. In the more realistic case where the moment conditions are correlated, the weighting matrix efficiently combines the moment conditions by accounting for different variances and nonzero correlations.

Several econometrics packages, including EViews, RATS and Stata, implement generalized method of moments fairly routinely. Why are Cragg-style estimators not used more? One problem is that the researcher must choose the additional moment conditions to be added in an ad hoc manner. Two researchers would generally use two different sets of moment conditions. Thus, the procedure would open one's research to the criticism of searching over different sets of moment conditions until the desired result is achieved. In fact, in large samples, one can improve on (or at least do no worse than) a previous researcher's estimator by adding more moment conditions. Where would one stop? A second issue is that ordinary least squares will be unbiased as well as consistent, whereas GMM is

⁴ The improvements over ordinary least squares estimation of the original model do not come by adding nonlinear functions of the x_j as independent variables and estimating the expanded equation by ordinary least squares or weighted least squares. The covariates appearing in the model do not change when we estimate the model by GMM. We simply add more zero correlation assumptions between the original error term and additional functions of the original covariates. These extra moment conditions take the form $E[f_h(x)u] = 0$, where $f_h(x)$ denotes a nonlinear function of x_1, x_2, \dots, x_k .

⁵ For textbook treatments of the choice of weighting matrix, see Hamilton (1994), Newey and McFadden (1994), Hayashi (2000), Ruud (2000) and Wooldridge (2001). If u is homoskedastic in the model, then the ordinary least squares estimator is just as efficient as Cragg's estimator.

⁶ For the problem of estimating the population mean when the variance is three times the mean, the two moment conditions are uncorrelated whenever y has a distribution symmetric about μ .

guaranteed only to be consistent. Generally, GMM can suffer from finite-sample problems, especially if one gets carried away and adds many moment conditions that do not add much information; for discussion, see Bound, Jaeger and Baker (1995), Altonji and Segal (1996) and Staiger and Stock (1997).

Given the additional decisions required in using estimators in the style of Cragg (1983) to improve on ordinary least squares results, it is little wonder that most applied researchers opt to stick with ordinary least squares. If they are concerned about heteroskedasticity, they have methods for computing standard errors and test statistics that are robust to heteroskedasticity of unknown form, as in White (1980). With large sample sizes, the additional efficiency gains that might be realized by Cragg's method probably pale in comparison to the questions raised by adding moment conditions to the ordinary least squares moment conditions. With small sample sizes, finite-sample bias in generalized method of moments estimators becomes an issue.

Whether one prefers ordinary least squares under the zero correlation assumption or a Cragg estimator under the zero conditional mean assumption, they share an important feature: each consistently estimates the parameters of interest without further distributional assumptions involving the error u . If possible, we want estimators to be robust to the failure of model assumptions that are not central to the problem at hand, like whether heteroskedasticity exists, or whether u has a particular distribution. Such robustness is the hallmark of method of moments estimation.⁷

Hansen's (1982) seminal work on generalized method of moments estimators demonstrated that moment conditions could be exploited very generally to estimate parameters consistently under weak assumptions. Hansen essentially showed that every previously suggested instrumental variables estimator, in linear or nonlinear models, with cross-section, time series or panel data, could be cast as a GMM estimator. Perhaps even more important, Hansen showed how to choose among the many possible method of moments estimators in a framework that allows for heteroskedasticity, serial correlation and nonlinearities.

As we saw with Cragg's (1983) estimator, an important feature of generalized method of moments is that it allows more moment conditions than there are parameters to estimate—that is, it allows the parameters to be *overidentified*. Generally, given the set of population moment conditions, an optimal weighting matrix can be obtained for a GMM analysis. (For details, see the textbook references given in note 5.)

One case where overidentification plays an important role is in the context of

⁷ This notion of robustness is distinct from insensitivity to data outliers, or fat-tailed distributions, which is often the kind of robustness that is the focus in the statistics literature. In our usage, the sample average is a robust estimator because it is consistent for the population mean whenever the population mean exists, regardless of distribution. But the sample average is very sensitive to outliers. Method of moments estimators are based on sample averages, and so they generally will be sensitive to outliers.

instrumental variables estimation.⁸ Suppose that one or more of the x_j in the population linear model is correlated with u , but that we have some variables properly excluded from the model that are uncorrelated with u . Provided the excluded variables are sufficiently correlated with the endogenous explanatory variables, we can use the excluded exogenous variables as instrumental variables.⁹ When we have more instrumental variables than are needed to estimate the parameters, the most common estimation method is two-stage least squares. The two-stage least squares estimator is a generalized method of moments estimator that uses a weighting matrix constructed under homoskedasticity. The optimal GMM estimator uses a weighting matrix identical to that described for Cragg's estimator.¹⁰ The optimal GMM estimator is asymptotically no less efficient than two-stage least squares under homoskedasticity, and GMM is generally better under heteroskedasticity.

A common theme about generalized method of moments is developing here. GMM estimators often can be found that are more efficient than common method of moments estimators—such as ordinary least squares and two-stage least squares—when textbook auxiliary assumptions such as homoskedasticity fail. Theoretically, this would seem to make a strong case for always using a GMM procedure. However, while virtually every empirical researcher has used ordinary least squares or two-stage least squares, most have probably never used a sophisticated method of moments estimator, which I take to be synonymous with GMM. In the next several sections, I discuss the scope of GMM for standard applications as well as for more sophisticated problems.

Cross-Section Applications

As mentioned in the previous section, it is always possible with overidentified parameters to improve upon two-stage least squares in the context of heteroskedasticity of unknown form. Still, application of generalized method of moments in place of two-stage least squares is rare for cross-section applications. The reason, I think, is that even when heteroskedasticity clearly exists, it often has only a minor impact on estimates of coefficients and statistical significance. Moreover, as is the case with ordinary least squares, there are methods for calculating standard errors for the two-stage least squares estimator that are robust to heteroskedasticity. The

⁸ Instrumental variables and two-stage least squares are discussed in the paper by Angrist and Krueger in this symposium.

⁹ In Cragg's (1983) estimator, the "instruments" are simply nonlinear functions of the original explanatory variables. The more common usage of instruments arises when some of the explanatory variables are endogenous, and then the instruments come from outside the equation of interest.

¹⁰ The precise definition of heteroskedasticity when u is the error and z denotes the entire set of exogenous variables is that $E(u^2|z)$ depends on z . White (1982) called the GMM estimator the two-stage instrumental variables estimator. If $E(u^2|z)$ is constant, two-stage least squares is an efficient GMM estimator.

additional gains from using GMM may be small. Besides, as practitioners know, the most important step in applying instrumental variable methods is finding good instruments. With poor instruments, the efficient GMM estimator is not likely to help much.

To illustrate the differences among ordinary least squares, two-stage least squares and efficient generalized method of moments in a cross-section setting, I use a subset of the data in Card (1995) to estimate the return to education for blacks and nonblacks. I use data on men who were living in the South in 1966; the wage data are for 1976. The structural model has $\log(\text{wage})$ as the dependent variable, where *wage* is measured hourly. The key explanatory variables are years of schooling (*educ*), a binary indicator for race (*black*) and an interaction between these two. The other explanatory variables include a quadratic in experience, an indicator for living in the South in 1976 and an indicator for living in an urban area in 1976.

The ordinary least squares estimates of the coefficients on *educ*, *black* and *black* · *educ* are given in the first column of Table 1. Both the usual and heteroskedasticity-robust standard errors are given (with the latter in brackets). The ordinary least squares estimate of the return to education for nonblacks is about 7.1 percent, and it is very statistically significant. The estimated return to education is about 1.1 percentage points higher for blacks, but the coefficient on the interaction term is insignificant at the 10 percent level.

A common concern in these sorts of wage regressions is that the education variable may be correlated with unobserved factors that can also affect earnings, such as motivation, ability or family background. As a result, the ordinary least squares estimator is generally biased and inconsistent for the causal effect of schooling on earnings. A standard solution to the endogeneity of education is to find an instrumental variable for education.

Following Card (1995), I use an indicator for whether the man lived near a four-year college at age 16 (called *nearc4*) as an instrument for education, and I use the interaction between *black* and *nearc4* as a natural instrument for *black* · *educ*. If I used only these two instrumental variables, the equation would be just identified, and there would be no difference between two-stage least squares and generalized method of moments.¹¹ Therefore, I add to the instrument list interactions between *nearc4* and the four other exogenous variables, resulting in four overidentifying restrictions. The two-stage least squares estimates are notably higher than the ordinary least squares estimates on *educ* and the interaction term. While *educ* becomes much less significant (but still significant at the 2 percent level), the interaction term *black* · *educ* becomes significant at the 0.1 percent level. The point

¹¹ When the number of instruments is the same as the number of explanatory variables, all general method of moments estimators, regardless of weighting matrix, reduce to the standard instrumental variables estimator. See Angrist and Krueger, this symposium, for more on the basic instrumental variables estimator.

Table 1
Estimates of a Wage Equation

<i>Explanatory Variables</i>	<i>Ordinary Least Squares</i>	<i>Two-Stage Least Squares</i>	<i>Generalized Method of Moments</i>
<i>educ</i>	.0708 (.0062) [.0066]	.1093 (.0463) [.0461]	.1062 (.0454)
<i>black</i>	-.3598 (.0972) [.0975]	-.6296 (.1446) [.1472]	-.6351 (.1452)
<i>black · educ</i>	.0114 (.0077) [.0078]	.0368 (.0110) [.0112]	.0371 (.0111)
Observations	1,247	1,247	1,247
R-Squared	.3038	.2521	N/A

Notes: The equations also contain an intercept along with a quadratic in potential experience and binary indicators for living in the South and living in an SMSA. The two-stage least squares and generalized method of moments estimates are obtained using all explanatory variables except *educ* and *black · educ* as instruments, and, in addition, *nearc4*, *black · nearc4* and interactions between *black* and the other four exogenous explanatory variables. For ordinary least squares and two-stage least squares, quantities in parentheses are the usual standard errors; those in brackets are robust to general heteroskedasticity.

estimate implies that another year of education is worth about 3.7 percentage points more for a black man than for a nonblack man.

The generalized method of moments estimator with weighting matrix that accounts for heteroskedasticity of unknown form uses the same list of instrumental variables as the two-stage least squares estimator.¹² As shown in Table 1, the GMM estimates and standard errors are very similar to those for two-stage least squares. Because the two sets of standard errors for the two-stage least squares estimates are very close, heteroskedasticity does not appear to be much of a problem. As a result, it is unsurprising that GMM and two-stage least squares yield similar results. One reaction to this example is that GMM provides no particular advantage over two-stage least squares. The point estimates and statistical significance are quite similar. A second reaction is that using GMM does not hurt anything, and perhaps with other models or data sets it might have offered greater precision. Both views seem sensible.

Recently, GMM has been applied successfully to estimate certain nonlinear models with endogenous explanatory variables that do not appear additively in an equation. A good example is an exponential regression function with endogenous explanatory variables. Mullahy (1997) uses GMM to estimate a model for daily

¹² EViews 4 was used to obtain the GMM estimates using the cross-section version of the weighting matrix to account for heteroskedasticity of unknown form. Implementing GMM is straightforward. Just as with two-stage least squares, one specifies the dependent variable, the explanatory variables and the list of all exogenous variables (including the instruments). The only thing different from two-stage least squares is in specifying the use of the efficient weighting matrix that accounts for possible heteroskedasticity.

cigarette consumption. Berry, Levinsohn and Pakes (1999) show how GMM can be used to estimate structural parameters to evaluate the welfare implications of voluntary export restrictions.¹³

Time Series Applications

Hansen (1982) introduced generalized method of moments estimation primarily with time series applications in mind, and so it is not surprising that GMM is relatively advantageous for time series data. In applications of linear time series models, serial correlation in the errors is the most important departure from common textbook assumptions. This raises the possibility of allowing the GMM weighting matrix to account for serial correlation of unknown form, as well as for heteroskedasticity, as discussed in Hansen (1982), White (1984) and Newey and West (1987).¹⁴

To obtain a more efficient estimator than two-stage least squares (or ordinary least squares), one must have overidentifying restrictions. With time series, we can add moment conditions by assuming that past values of explanatory variables, or even past values of the dependent variable, are uncorrelated with the error term, even though they do not appear in the model. The drawback to finding moment conditions in this way is that it restricts the dynamics in the model. For example, if we start with a static equation, such as a simple Phillips curve with inflation as the dependent variable, we could apply generalized method of moments using current and lagged unemployment as the instrumental variables. The weighting matrix would account for possible serial correlation in the errors (often interpreted as supply shocks). But such an approach would assume that current and past unemployment rates are uncorrelated with the supply shocks, something we may not want to assume.

Using lagged values of dependent and independent variables makes more sense in the context of models estimated under rational expectations. Then, the error term in the equation is uncorrelated with all variables dated at earlier time periods. Campbell and Mankiw (1990) use an instrumental variables approach to test the permanent income hypothesis. They also estimate the fraction of consumers who consume out of current, rather than permanent, income, by estimating a

¹³ Unfortunately, generalized method of moments does not help much in relaxing distributional assumptions for many important nonlinear models with endogenous explanatory variables, such as probit or Tobit. For other models, such as the Box-Cox regression model, parameters can be estimated consistently under weak distributional assumptions (Amemiya and Powell, 1981), but then no partial effects on the mean or median value of the response variable can be estimated without making stronger assumptions. Just because one can estimate parameters by GMM under weak assumptions does not necessarily mean that quantities of interest can be estimated under those same assumptions.

¹⁴ Importantly, solving serial correlation problems in u_t by feasible generalized least squares, such as Cochrane-Orcutt, generally requires certain strict exogeneity assumptions on the regressors; that is, we must assume that the covariates in some time periods other than t are uncorrelated with u_t (Wooldridge, 2000, chapter 12). These assumptions are often questionable in time series contexts.

simple equation relating consumption changes to income changes. For their instrumental variables for income changes, Campbell and Mankiw use lags of consumption and income changes dated two or more quarters in the past, as they are worried about time aggregation of the consumption and income data inducing serial correlation in the errors. Campbell and Mankiw account for serial correlation by using two-stage least squares and computing robust standard errors. A similar problem is faced by Clarida, Gali and Gertler (2000), who estimate forward-looking policy reaction functions for the Federal Reserve. In an equation with the federal funds interest rate as the dependent variable, Clarida, Gali and Gertler note that the errors will follow a moving average process if the Fed's target horizon for inflation or the output gap exceeds the frequency of the quarterly data. These authors implement GMM, where the weighting matrix accounts for the serial correlation (and possible heteroskedasticity). Moving average error processes also arise in estimating asset pricing models where the investment horizon differs from the data frequency. In fact, such situations were an important motivation for Hansen (1982); see Hansen and Hodrick (1980) for an empirical application.

Hansen and Singleton (1982) apply generalized method of moments to estimate nonlinear asset pricing models under rational expectations. The moment conditions used by Hansen and Singleton (1982) depend nonlinearly on two economic parameters, the discount rate and the coefficient of relative risk aversion. GMM has significant advantages over maximum likelihood in this context because GMM allows estimation under the restrictions implied by the theory; there is no need to add distributional assumptions that are not implied by the theory.¹⁵ Hansen and Singleton (1982) obtain estimates of the discount rate close to, but always less than, unity and a coefficient of relative risk aversion ranging from about 0.35 to 0.97, depending on the kind and number of returns used and the set of instruments used in GMM estimation. Recently, Stock and Wright (2000), Weber (2000) and Neely, Roy and Whiteman (2001) offer explanations for the wide disparities in estimates of consumption-based asset pricing models.

Panel Data Applications

Some of the most interesting recent applications of generalized method of moments are to panel data. I will focus attention on linear unobserved effects models where the unobserved effect, or unobserved heterogeneity, is allowed to be correlated with the observed covariates. The standard estimator used to eliminate the potential bias caused by omitted heterogeneity is the fixed effects, or within,

¹⁵ Generally, there are no computationally simple alternatives to GMM in nonlinear models, as the sample moment conditions are nonlinear in the parameters, and so any estimation method requires iterative methods.

estimator.¹⁶ The fixed effects estimator, which is a method of moments estimator based on the data after subtracting off time averages, is popular because it is simple, easily understood, and robust standard errors are readily available (for example, Wooldridge, 2001, chapter 10).

When analyzing the fixed effects estimator, the standard assumptions are that the time-varying errors have zero means, constant variances and zero correlations, all conditional on the observed history of the covariates and on the unobserved effect (for example, Wooldridge, 2001, chapter 10). The first assumption, that the conditional mean of the time-varying errors is zero, implies that the observed covariates in every time period are uncorrelated with the time-varying errors in each time period. This so-called strict exogeneity assumption for the covariates is crucial for consistency of the fixed effects estimator. But the assumptions about constant variance and no serial correlation are used primarily to simplify calculation of standard errors. If either heteroskedasticity or serial correlation is present, a generalized method of moments procedure can be more efficient than the fixed effects estimator, although the likely gains in standard applications are largely unknown. Extra moment conditions are available from the assumption that the covariates in all time periods are assumed to be uncorrelated with each time-varying error. Wooldridge (2001, chapter 11) describes how to implement GMM in this case.¹⁷

Generalized method of moments is convenient for estimating interesting extensions of the basic unobserved effects model, for example, models where unobserved heterogeneity interacts with observed covariates. Lemieux (1998) uses GMM to estimate the union-wage effect when unobserved heterogeneity is valued differently in the union and nonunion sectors.

Generalized method of moments is applied more often to unobserved effects models when the explanatory variables are not strictly exogenous even after controlling for an unobserved effect. As in cross-section and time series cases, there is usually a convenient estimator that is consistent quite generally, but possibly inefficient relative to GMM. For example, for studying the effects of prison population on crime rates, Levitt (1996) uses pooled two-stage least squares on a panel data set of states, after removing state fixed effects by differencing adjacent years. If the errors in the first-differenced equation are homoskedastic and serially uncorrelated, the pooled two-stage least squares estimator is efficient. If not, a GMM estimator can improve upon two-stage least squares.

Another leading application of generalized method of moments in panel data contexts is when a model contains a lagged dependent variable along with an unobserved effect. The standard method of estimating such models dates back to Anderson and Hsiao (1982): first-differencing is used to eliminate the unobserved

¹⁶ Differencing across different time periods is another common method of eliminating the unobserved effects.

¹⁷ Chamberlain (1984) describes minimum distance estimation of unobserved effects panel data models. It turns out that the minimum distance and GMM approaches are asymptotically equivalent.

effect, and then lags two and beyond are used as instrumental variables for the differenced lagged dependent variable. Because the original time-varying errors are assumed to be serially uncorrelated, the differenced errors must contain serial correlation. GMM is well suited for obtaining efficient estimators that account for the serial correlation; see, for example, Arellano and Bond (1991). Van Reenen (1996) applies GMM to estimate a dynamic wage equation that measures the amount of firm rents captured by workers. For estimating a dynamic labor demand model using firm-level data, Blundell and Bond (1998) find that GMM with additional moment conditions can provide more precise estimates than can two-stage least squares of the parameter on lagged labor demand.

Concluding Remarks

The method of moments can be used to obtain parameter estimators that are consistent under weak distributional assumptions. In standard settings, where one would typically use ordinary or two-stage least squares, or standard panel data methods such as fixed effects, generalized method of moments can be used to improve over the standard estimators when auxiliary assumptions fail, at least in large samples. However, because basic econometric methods can be used with robust inference techniques that allow for arbitrary heteroskedasticity or serial correlation, the gains to practitioners from using GMM may be small. Significant GMM improvements are most likely in time series or panel data applications with neglected serial correlation. GMM is indispensable for more sophisticated applications, including nonlinear rational expectations models or dynamic unobserved effects panel data models.

■ *I would like to thank Gary Chamberlain, Bradford De Long, Alan Krueger and Michael Waldman for very helpful comments on an earlier draft. Timothy Taylor's editorial advice vastly improved the organization and content of the paper.*

References

- Altonji, Joseph G. and Lewis M. Segal.** 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures." *Journal of Business and Economic Statistics*. July, 14:3, pp. 353-66.
- Amemiya, Takeshi and James L. Powell.** 1981. "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-Linear Two Stage Least Squares Estimator." *Journal of Econometrics*. December, 17:3, pp. 351-81.
- Anderson, T. W. and Cheng Hsiao.** 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics*. January, 18:1, pp. 47-82.
- Arellano, Manuel and Stephen R. Bond.** 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*. April, 58:2, pp. 277-97.
- Berry, Steven, James Levinsohn and Ariel Pakes.** 1999. "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy." *American Economic Review*. June, 89:3, pp. 400-30.
- Blundell, Richard and Stephen Bond.** 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*. November, 87:1, pp. 115-43.
- Bound, John, David A. Jaeger and Regina M. Baker.** 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*. June, 90:430, pp. 443-50.
- Campbell, John Y. and N. Gregory Mankiw.** 1990. "Permanent Income, Current Income, and Consumption." *Journal of Business and Economic Statistics*. July, 8:3, pp. 265-79.
- Card, David.** 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behavior: Essays in Honor of John Vanderkamp*. L. N. Christophides, E. K. Grant and R. Swidinsky, eds. Toronto: University of Toronto Press, pp. 201-22.
- Chamberlain, Gary.** 1984. "Panel Data," in *Handbook of Econometrics, Volume Two*. Z. Griliches and M.D. Intriligator, eds. Amsterdam: North-Holland, pp. 1248-318.
- Clarida, Richard, Jordi Gali and Mark Gertler.** 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics*. February, 115:1, pp. 147-80.
- Cragg, John G.** 1983. "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form." *Econometrica*. May, 51:3, pp. 751-63.
- Hamilton, James D.** 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Hansen, Lars Peter.** 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*. July, 50:4, pp. 1029-054.
- Hansen, Lars Peter and Robert J. Hodrick.** 1980. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." *Journal of Political Economy*. October, 88:5, pp. 829-53.
- Hansen, Lars Peter and Kenneth J. Singleton.** 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica*. September, 50:5, pp. 1269-286.
- Hayashi, Fumio.** 2000. *Econometrics*. Princeton, N.J.: Princeton University Press.
- Lemieux, Thomas.** 1998. "Estimating the Effects of Union on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection." *Journal of Labor Economics*. April, 16:2, pp. 261-91.
- Levitt, Steven D.** 1996. "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation." *Quarterly Journal of Economics*. May, 111:2, pp. 319-51.
- Mullahy, John.** 1997. "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior." *Review of Economics and Statistics*. November, 79:4, pp. 586-93.
- Neely, Christopher J., Amlan Roy and Charles H. Whiteman.** 2001. "Risk Aversion Versus Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model." *Journal of Business and Economic Statistics*. Forthcoming.
- Newey, Whitney K. and Daniel L. McFadden.** 1994. "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Volume Four*. R.F. Engle and D.L. McFadden, eds. Amsterdam: North-Holland, pp. 2111-245.
- Newey, Whitney K. and Kenneth D. West.** 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*. May, 55:3, pp. 703-08.
- Ruud, Paul A.** 2000. *An Introduction to Classical Econometric Theory*. Oxford: Oxford University Press.
- Staiger, Douglas and James H. Stock.** 1997.

"Instrumental Variables Regression with Weak Instruments." *Econometrica*. May, 65:3, pp. 557–86.

Stigler, Stephen M. 1986. *The History of Statistics*. Cambridge, Mass.: Harvard University Press.

Stock, James H. and Jonathan H. Wright. 2000. "GMM with Weak Identification." *Econometrica*. September, 68:5, pp. 1055–096.

Van Reenen, John. 1996. "The Creation and Capture of Rents: Wages and Innovations in a Panel of U.K. Companies." *Quarterly Journal of Economics*. February, 111:1, pp. 195–226.

Weber, Christian E. 2000. "'Rule-of-Thumb' Consumption, Intertemporal Substitution, and Risk Aversion." *Journal of Business and Economic Statistics*. October, 18:4, pp. 497–502.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*. May, 48:4, pp. 817–38.

White, Halbert. 1982. "Instrumental Variables Regression with Independent Observations." *Econometrica*. March, 50:2, pp. 483–99.

White, Halbert. 1984. *Asymptotic Theory for Econometricians*. Orlando, Fla.: Academic Press.

Wooldridge, Jeffrey M. 2000. *Introductory Econometrics: A Modern Approach*. Cincinnati, Ohio: South-Western.

Wooldridge, Jeffrey M. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.