GENERALIZED METHOD OF MOMENTS ESTIMATION

Econometrics C ♦ Lecture Note 8
Heino Bohn Nielsen
September 23, 2016

GMM estimation is an alternative to the likelihood principle and it has been widely used the last 20 years. This note introduces the principle of GMM estimation and discusses some familiar estimators, OLS, IV, 2SLS and ML, as special cases. We focus on the intuition for the procedure, but GMM estimation is inherently technical and some details are discussed along the way. The note first presents the general theory. It then considers the special case of linear instrumental variables estimation and derives the well-known IV estimators as special cases. Finally, we present a GMM module for OxMetrics and discuss two empirical examples. One is the estimation of monetary policy rules. The other is the estimation of Euler equations.

Outline

1 Introduction

Consider first the linear regression model as given by

$$y_t = x_t'\beta_0 + \epsilon_t, \quad t = 1, 2, ..., T,$$

where y_t is a scalar variable to be explained and x_t is a $K-$dimensional vector of regressors. The parameter in the model is here called β and we use β_0 to denote the true value of the parameter β. Earlier in this course (see Lecture Note 2), the OLS estimator was derived from the assumption of a zero-conditional mean, $E[\epsilon_t \mid x_t] = 0$, which allowed a natural interpretation of the parameter,

$$\beta_0 = \frac{\partial E[y_t \mid x_t]}{\partial x_t} = \frac{\partial\,(x_t'\beta_0)}{\partial x_t}.$$

This assumption implies the (unconditional) *moment condition*,

$$E[x_t\epsilon_t] = E[x_t(y_t - x_t'\beta_0)] = 0, \tag{1}$$

and if the data satisfies conditions for a law of large numbers, the OLS estimator can be derived from the sample-counterpart to (1), i.e.

$$\frac{1}{T}\sum_{t=1}^{T} x_t(y_t - x_t'\hat{\beta}) = 0,$$

and we referred to $\hat{\beta}$ as the method-of-moment estimator.

If the model of interest is *non-linear*, e.g.

$$y_t = h(x_t, \phi_0) + \epsilon_t,$$

where $h(x_t, \phi_0) = E[y_t \mid x_t]$ is some non-linear function, e.g.

$$h(x_t, \phi_0) = \exp(x_t'\phi_0),$$

then the zero conditional mean requirement would be the same, $E[\epsilon_t \mid x_t] = 0$, and estimation could be based on the moment condition

$$E[x_t\epsilon_t] = E[x_t(y_t - h(x_t, \phi_0)] = E[x_t(y_t - \exp(x_t'\phi_0)] = 0, \tag{2}$$

and the sample counterpart

$$\frac{1}{T}\sum_{t=1}^{T} x_t(y_t - \exp(x_t'\hat{\phi})) = 0.$$

We could solve the sample moment conditions (numerically) to obtain $\hat{\phi}$.

You may also recall the instrumental variables estimator in the linear model, introduced in an earlier course. This estimator was based on the existence of $R \geq K$ instrumental variables, z_t say, that are uncorrelated with ϵ_t, i.e.

$$E[z_t\epsilon_t] = E[z_t(y_t - x_t'\beta_0)] = 0, \tag{3}$$

and the IV estimator was found from the sample-counterpart to (3).

Observe that the moment conditions in (1), (2), and (3) have the same form and the OLS estimator, the non-linear estimator, the IV estimator, and the two-stage least squares estimator are all examples of the same general estimation principle, known as the *generalized method of moments* (GMM). This note introduces the GMM principle and complements the coverage in Verbeek (2004, Section 5.6). The idea of GMM is intuitive as well as elegant, and knowledge of the principle of GMM is very useful for the understanding of econometrics in general.

Compared to the maximum likelihood (ML) principle applied earlier in the course, GMM estimation is based on fewer assumptions. To formulate a likelihood function for the non-linear regression model, for example, we need to specify the entire distribution for $y_t \mid x_t$, while GMM is based solely on the moment condition in (2). On the one hand, this makes the model formulation easier, because we do not care about the distribution of ϵ_t but only the exogenenity (or pre-determinedness) of the variables. On the other hand it may make the model control more complicated, because the sign of misspecification is less obvious. In addition, the GMM estimator may be less efficient (i.e. have a larger asymptotic variance) than the ML estimator, because it is based on less *a priori* information.

1.1 Econometrics for Rational Expectations Models

One particularly interesting case, where the GMM estimator is frequently used, is the case of models derived from economic theory under the assumption of rational expectations. Economic models with rational expectations typically imply that agents do not make systematic errors, such that if they want to predict some variable, e.g. the function $u(w_{t+1}, \theta_0)$, based on the available information set, $\mathcal{I}_t$, it holds that

$$E[u(w_{t+1}, \theta_0) \mid \mathcal{I}_t] = 0. \tag{4}$$

The function $u(w_{t+1}, \theta_0)$ is a (potentially non-linear) function of future observations, w_{t+1}, involving the parameter θ_0, and it could represent an Euler equation for the optimal consumption decision or a condition for optimal monetary policy–we will see examples of this below.

For a vector of variables contained in the information set, $z_t \in \mathcal{I}_t$, the condition in (4) implies the unconditional expectation

$$E[u(w_{t+1}, \theta_0) \cdot z_t] = 0, \tag{5}$$

which is a moment condition, similar to (1), (2), or (3), stating that the variables z_t are uncorrelated with $u(w_{t+1}, \theta_0)$. Using the principle of GMM, this is typically sufficient to derive a consistent estimator, $\widehat{\theta}$.

GMM estimation is often closely connected to economic theory–exploiting directly a moment condition like (5). Consistency of GMM requires that the moment conditions (and hence the economic theories) are true. So whereas the imposed *statistical assumptions* are very mild, the GMM estimator is typically derived under very strict *economic assumptions*: for example a representative agent, global optimization, rational expectations etc.

For empirical applications there is typically an important difference in the approach of a likelihood analysis and a GMM estimation. The likelihood analysis begins with a statistical description of the data, and the econometrician should ensure that the likelihood function accounts for the main characteristics of the data. Based on the likelihood function we can test hypotheses implied by economic theory. A GMM estimation, on the other hand, typically begins with an economic theory and the data are used to produce estimates of the model parameters. Estimation is done under minimal statistical assumptions, and often less attention is given to the fit of the model.

1.2 Outline of the Note

The GMM principle is very general, and many known estimators can be seen as special cases of GMM. This means that GMM can be used as a unifying framework to explain the properties of estimators. Below we consider ordinary least squares (OLS), instrumental variables (IV), and ML estimators as special cases, and we derive the properties of the estimators under minimal assumptions. As an example, we characterize the properties of the ML estimator if the likelihood function is misspecified, see Box 2 below.

The rest of the note is organized as follows. In §2 we show how moment conditions can be used for estimation and we present the general theory for GMM estimation. In §3 we look at a particular class of GMM problems known as the instrumental variables estimation and we cover the well-known linear instrumental variables case in §3.1. In §4 we present some empirical examples: Most importantly forward looking monetary policy rules for the US, and nonlinear Euler equations for intertemporal optimization.

2 Moment Conditions and GMM Estimation

In this section we introduce the concept of a moment condition and discuss how moment conditions can be used for estimating the parameters of an econometric model. We then give the general formulation of GMM and outline the main properties.

2.1 Moment Conditions and Method of Moments (MM) Estimation

A *moment condition* is a statement involving the data and the parameters of interest. We use the general formulation

$$g(\theta_0) = E[f(w_t, z_t, \theta_0)] = 0, \tag{6}$$

where θ is a $K \times 1$ dimensional vector of parameters with true value θ_0; $f(\cdot)$ is an R dimensional vector of potentially non-linear functions; w_t is a vector of *model variables* that appear in the equation to be estimated; and z_t is a vector of so-called *instruments*.

In most applications the distinction between model variables (w_t) and instruments (z_t) is clear. If not we can define $f(y_t, \theta_0)$ where y_t includes all the observed data. The difference is discussed in more details in §3. The R equations in (6) simply state that the expectation of the function $f(w_t, z_t, \theta)$ is zero if evaluated at the true value θ_0. Observe that the moment conditions in the initial examples, (1), (2), and (3), are all specific examples of the general formulation in (6) for different choices of variables and different functions $f(\cdot)$.

If we knew the expectations then we could solve the equations in (6) to find θ_0, and for the system to be well-defined the solution should be unique. The presence of a unique solution is called *identification*:

Definition 1 (identification): *The moment conditions in (6) are said to identify the parameters in θ_0 if there is a unique solution, so that $E[f(w_t, z_t, \theta)] = 0$ if and only if $\theta = \theta_0$.*

For a given set of observations, w_t and z_t ($t = 1, 2, ..., T$), we cannot calculate the expectation, and it is natural to rely on sample averages. We define the analogous *sample moments* as

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \theta), \tag{7}$$

which contains the basic information in the data. Observe that because of sample uncertainty in the R sample moments, we have in general that

$$g_T(\theta_0) = \frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \theta_0) \neq 0,$$

and we cannot find the true value of the parameter. Instead we find an estimator, $\widehat{\theta}$, based on the sample moments. If $R = K$ we say that the system is *exactly identified*, and we solve to R equations with K unknown,

$$g_T(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \hat{\theta}) = 0,$$

and the resulting estimator is referred to as the *method of moments* (MM) estimator, compare Lecture Note 2.

If $R > K$ we have more equations than unknown parameters, and there is no solution to $g_T(\theta) = 0$. Instead we minimize a weighted sum of squares to find the GMM estimator, see Section 2.2 below.

If $R < K$ we have fewer equations than unknowns and solutions are not unique. In this case the parameter is not identified, and $R \geq K$ is known as the *order condition* for identification. To illustrate, consider the following four examples.

Example 1 (mm estimator of the mean): Suppose that y_t is a random variable drawn from a population with expectation μ_0, such that

$$g(\mu_0) = E[f(y_t, \mu_0)] = E[y_t - \mu_0] = 0,$$

where

$$f(y_t, \mu_0) = y_t - \mu_0.$$

Based on an observed sample, y_t $(t = 1, 2, ..., T)$, we can construct the corresponding sample moment conditions by replacing the expectation with the sample average:

$$g_T(\widehat{\mu}) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{\mu}) = 0. \tag{8}$$

This is one equation with one unknown, and the MM estimator of the mean μ_0 is the solution to (8), i.e.

$$\widehat{\mu}_{MM} = T^{-1}\sum_{t=1}^{T} y_t.$$

Note that the MM estimator is the sample average of y_t. ♦

Example 2 (ols as an mm estimator): Consider the linear regression model

$$y_t = x_t'\beta_0 + \epsilon_t, \quad t = 1, 2, ..., T, \tag{9}$$

where x_t is $K \times 1$ vector of regressors, and assume that it represents the conditional expectation: $E[y_t \mid x_t] = x_t'\beta_0$, such that $E[\epsilon_t \mid x_t] = 0$. This implies the K unconditional moment conditions

$$g(\beta_0) = E[x_t\epsilon_t] = E\left[x_t\left(y_t - x_t'\beta_0\right)\right] = 0.$$

Defining the corresponding sample moment conditions,

$$g_T(\widehat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} x_t\left(y_t - x_t'\widehat{\beta}\right) = \frac{1}{T}\sum_{t=1}^{T} x_t y_t - \frac{1}{T}\sum_{t=1}^{T} x_t x_t'\widehat{\beta} = 0,$$

we have K equations with K unknowns, and the MM estimator can be derived as the unique solution:

$$\widehat{\beta}_{MM} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t y_t\right), \tag{10}$$

provided that $\frac{1}{T}\sum_{t=1}^{T} x_t x_t'$ is non-singular such that the inverse exists. We recognize (10) as the OLS estimator, and recall the two conditions for identification: The moment conditions implied by predetermined regressors and the non-singularity implied by no-perfect-multicollinearity. ♦

Example 3 (under-identification and non-consistency): Now we reconsider the estimation model in equation (9) but we assume that some of the variables in x_t are endogenous in the sense that they are correlated with the error term. In particular, we write the partitioned regression model:

$$y_t = x_{1t}'\gamma_0 + x_{2t}'\delta_0 + \epsilon_t,$$

where the K_1 variables in x_{1t} are *predetermined*, while the $K_2 = K - K_1$ variables in x_{2t} are *endogenous*, i.e.

$$E[x_{1t}\epsilon_t] \quad = \quad 0 \quad (K_1 \times 1) \tag{11}$$

$$E[x_{2t}\epsilon_t] \quad \neq \quad 0 \quad (K_2 \times 1). \tag{12}$$

In this case OLS is known to be inconsistent. As a MM estimator, the explanation is that we have K parameters in $\beta_0 = (\gamma_0', \delta_0')'$, but only $K_1 < K$ moment conditions. The K_1 equations with K unknowns have no unique solution, so the parameters are not identified by the model. ♦

Example 4 (simple iv estimator): Consider the estimation problem in Example 3, but now assume that there exist K_2 new variables, z_{2t}, that are correlated with x_{2t} but uncorrelated with the errors:

$$E[z_{2t}\epsilon_t] = 0. \tag{13}$$

The K_2 new moment conditions in (13) can replace (12). To simplify notation, we define

$$\underset{(K\times 1)}{x_t} = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} \quad \text{and} \quad \underset{(K\times 1)}{z_t} = \begin{pmatrix} x_{1t} \\ z_{2t} \end{pmatrix},$$

where x_t are the *model variables* and z_t are the *instruments*. We say that the predetermined variables are instrument for themselves, while the *new instruments*, z_{2t}, are instruments for x_{2t}. Using (11) and (13) we have K moment conditions:

$$g(\beta_0) = E[z_t\epsilon_t] = E[z_t\left(y_t - x_t'\beta_0\right)] = 0.$$

The corresponding sample moment conditions are given by

$$g_T(\widehat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} z_t\left(y_t - x_t'\widehat{\beta}\right) = 0,$$

and the MM estimator is the unique solution:

$$\widehat{\beta}_{MM} = \left(\frac{1}{T}\sum_{t=1}^{T} z_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} z_t y_t\right),$$

provided that the $K \times K$ matrix $\sum_{t=1}^{T} z_t x_t'$ can be inverted. This is the case if the new instruments are correlated with the endogenous variables; we say the instruments are *relevant*. Observed, that this MM estimator coincides with the simple IV estimator. ♦

2.2 Generalized Method of Moments (GMM) Estimation

The case $R > K$ is referred to as *over-identification* and the estimator is denoted the GMM estimator. In this case there are more equations than parameters and no solution to

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta) = 0$$

in general.

Instead we find an estimator by minimizing the distance from the vector $g_T(\theta)$ to zero. One possibility is to choose θ to minimize the simple distance corresponding to the sum of squares, $g_T(\theta)' g_T(\theta)$. That has the disadvantage of being dependent on the scaling of the moments (e.g. whether a price index is scaled so that $1980 = 100$ or $1980 = 1$), and more generally we could minimize the weighted sum of squares, defined by the quadratic form

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta), \tag{14}$$

where W_T is an $R \times R$ symmetric and positive definite *weight matrix* that attach weights to the individual moments. We can think of the matrix W_T as a weight matrix reflecting the importance of the moments; alternatively we can think of W_T as defining the metric for measuring the distance from $g_T(\theta)$ and zero. Note that the GMM estimator depends on the chosen weight matrix:

$$\widehat{\theta}_{GMM}(W_T) = \arg\ \min_{\theta}\ \left\{ g_T(\theta)' W_T g_T(\theta) \right\}. \tag{15}$$

Since (14) is a quadratic form it holds that $Q_T(\theta) \geq 0$. Equality holds for the exactly identified case, where the weight matrix is redundant and the estimator $\widehat{\theta}_{MM}$ unique.

To derive the estimator in (15) we take the first derivative and solve the K equations

$$\frac{\partial Q_T(\theta)}{\partial \theta} = \underset{(K\times 1)}{0},$$

for the K unknown parameters in θ. In some cases these equations can be solved analytically to produce the GMM estimator, $\widehat{\theta}_{GMM}$, and we will see one example from a linear model below. If the function $f(w_t, z_t, \theta)$ is non-linear, however, it is in most cases not possible to find an analytical solution, and we have to rely on a numerical procedure for minimizing $Q_T(\theta)$.

2.3 Properties of the GMM Estimator

To obtain consistency of the GMM estimator we need a law of large numbers to apply. We therefore make the following assumption.

Assumption 1 (law of large numbers): *The data are such that a law of large numbers applies to $f(w_t, z_t, \theta)$, i.e. $T^{-1}\sum_{t=1}^{T} f(w_t, z_t, \theta) \to E[f(w_t, z_t, \theta)]$ for $T \to \infty$.*

For simplicity the assumption is formulated directly on $f(w_t, z_t, \theta)$, but it is a restriction on the behavior of the data and the assumption can be translated into precise requirements. For IID data the assumption is fulfilled, while for time series we require stationarity and weak dependence known from OLS regression.

RESULT 1 (CONSISTENCY): *Let the data obey Assumption 1. If the moment conditions are correct, $g(\theta_0) = 0$, then (under some regularity conditions): $\widehat{\theta}_{GMM}(W_T) \to \theta_0$ as $T \to \infty$ for all W_T positive definite.*

Different weight matrices produce different estimators, and Result 1 states that although they may differ for a given data set they are all consistent! The intuition is the following: If a law of large numbers applies to $f(w_t, z_t, \theta)$, then the sample moment, $g_T(\theta)$, converges to the population moment, $g(\theta)$. And since $\widehat{\theta}_{GMM}(W_T)$ makes $g_T(\theta)$ as close a possible to zero, it will be a consistent estimator of the solution to $g(\theta_0) = 0$. The requirement is that W_T is positive definite, so that we put a positive and non-zero weight on all moment conditions. Otherwise we may throw important information away.

To derive the asymptotic distribution of the estimator we assume that a central limit theorem holds for $f(w_t, z_t, \theta)$. In particular, we assume the following:

ASSUMPTION 2 (CENTRAL LIMIT THEOREM): *The data are such that a central limit theorem applies to $f(w_t, z_t, \theta)$:*

$$\sqrt{T} \cdot g_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(w_t, z_t, \theta_0) \to N(0, S), \tag{16}$$

where S is the asymptotic variance of $f(w_t, z_t, \theta_0)$.

This is again a high-level assumption that translates into requirements on the data. It is beyond the scope of this note to state the precise requirements, but they are similar to the requirements needed for deriving the distribution of the OLS estimator. The requirements for Assumption 2 are typically stronger than for Assumption 1. As we shall see below, the asymptotic variance matrix, S, plays an important role in GMM estimation; and many of the technicalities of GMM are related to the estimation of S.

RESULT 2 (ASYMPTOTIC DISTRIBUTION OF GMM): *Let the data obey Assumptions 1 and 2. For a positive definite weight matrix W, the asymptotic distribution of the GMM estimator is given by*

$$\sqrt{T}\left(\widehat{\theta}_{GMM} - \theta_0\right) \to N(0, V). \tag{17}$$

The asymptotic variance is given by

$$V = \left(D'WD\right)^{-1} D'WSWD \left(D'WD\right)^{-1}, \tag{18}$$

where

$$D = E\left[\frac{\partial f(w_t, z_t, \theta)}{\partial \theta'}\right]$$

is the expected value of the $R \times K$ matrix of first derivatives of the moment function $f(\cdot)$.

A sketch of the derivation of the asymptotic distribution of the GMM estimator is given in Box 1. This Box is not a part of the formal curriculum, but it illustrates the statistical tools. The expression for the asymptotic variance in (18) is quite complicated. It depends on S, the chosen weight matrix, W, and the expected derivative D. For the latter, you could think of the derivative of the sample moments

$$D_T = \frac{\partial g_T(\theta)}{\partial \theta'} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial f(w_t, z_t, \theta)}{\partial \theta'}, \tag{19}$$

and D is the limit of D_T for $T \to \infty$.

2.4 Efficient GMM Estimation

It follows from Result 2 that the variance of the estimator depends on the weight matrix, W_T; some weight matrices may produce precise estimators while other weight matrices produce poor estimators with large variances. We want to find a systematic way of choosing the good estimators. In particular we want to select a weight matrix, W_T^{opt}, that produces the estimator with the smallest possible asymptotic variance. This estimator is denoted the *efficient–* or *optimal GMM estimator*.

It seems intuitive that moments with a small variance are very informative on the parameters and should have a large weight while moments with a high variance should have a smaller weight. And it can be shown that the *optimal weight matrix*, W_T^{opt}, has the property that

$$\text{plim}\, W_T^{opt} = S^{-1}.$$

With an optimal weight matrix, $W = S^{-1}$, the asymptotic variance in (17) simplifies to

$$V = \left(D'S^{-1}D\right)^{-1} D'S^{-1}SS^{-1}D \left(D'S^{-1}D\right)^{-1} = \left(D'S^{-1}D\right)^{-1}, \tag{20}$$

which is the smallest possible asymptotic variance.

Result 3 (asymptotic distribution of efficient gmm): *The asymptotic distribution of the efficient GMM estimator is given in (17), with asymptotic variance (20).*

To interpret the asymptotic variance in (20), we note that the best moment conditions are those for which S is small and D is large (in a matrix sense). A small S means that the sample variation of the moment (or the noise) is small. D is the derivative of the moment, so a large D means that the moment condition is much violated if $\theta \neq \theta_0$, and the moment is very informative on the true values, θ_0. This is also related to the curvature of the criteria function, $Q_T(\theta)$, similar to the interpretation of the expression for the variance of the ML estimator.

Box 1: Asymptotic Properties of the GMM Estimator

For a large sample, the consistent estimator, $\widehat{\theta} = \widehat{\theta}_{GMM}$, is close to the true value, θ_0. To derive the asymptotic distribution we use a first order Taylor approximation of $g_T(\theta)$ around the true value θ_0 to obtain

$$g_T(\theta) \approx g_T(\theta_0) + D_T\,(\theta - \theta_0)\,, \tag{B1-1}$$

where $D_T = \partial g_T(\theta)/\partial\theta'$ is the $R \times K$ matrix of first derivatives. Inserting (B1-1) in the criteria function (14) yields

$$\begin{aligned} Q_T(\theta) &\approx (g_T(\theta_0) + D_T\,(\theta - \theta_0))'\,W\,(g_T(\theta_0) + D_T\,(\theta - \theta_0)) \\ &= g_T(\theta_0)'Wg_T(\theta_0) + g_T(\theta_0)'WD_T\,(\theta - \theta_0) \\ &\quad + (\theta - \theta_0)'\,D_T'Wg_T(\theta_0) + (\theta - \theta_0)'\,D_T'WD_T\,(\theta - \theta_0)\,. \end{aligned}$$

To minimize the criteria function we find the first derivative,

$$\frac{\partial Q_T(\theta)}{\partial\theta} = g_T(\theta_0)'WD_T + D_T'Wg_T(\theta_0) + 2 \cdot D_T'WD_T\,(\theta - \theta_0)\,.$$

Noting that the first two terms are scalars, $g_T(\theta_0)'WD_T = D_T'Wg_T(\theta_0)$, the first order condition for a minimum is given by

$$D_T'Wg_T(\theta_0) + D_T'WD_T\left(\widehat{\theta} - \theta_0\right) = 0.$$

By collecting terms we get

$$\widehat{\theta} = \theta_0 - \left(D_T'WD_T\right)^{-1} D_T'Wg_T(\theta_0), \tag{B1-2}$$

which expresses the estimator as the true value plus an estimation error.

To discuss the asymptotic behavior we define the limit

$$D = \operatorname{plim} D_T = E\left[\frac{\partial f(w_t, z_t, \theta)}{\partial\theta'}\right].$$

Consistency then follows from

$$\operatorname{plim}\widehat{\theta} = \theta_0 - \left(D'WD\right)^{-1} D'Wg(\theta_0) = \theta_0,$$

where we have used that $g(\theta_0) = 0$.

To derive the asymptotic distribution we recall that $\sqrt{T} \cdot g_T(\theta_0) \to N(0, S)$. It follows directly from (B1-2) that the asymptotic distribution of the estimator is given by

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \to N(0, V),$$

where the asymptotic variance is

$$V = \left(D'WD\right)^{-1} D'WSWD\left(D'WD\right)^{-1}.$$

Box 2: Pseudo-Maximum-Likelihood (PML) Estimation

The asymptotic properties of the maximum likelihood estimator is typically derived under the assumption that the likelihood function is correctly specified. That is a strong assumption, which may not be fulfilled in all applications. In this box we illustrate that the ML estimator can also be obtained as the solution to a set of moment conditions. An estimator which is derived from maximizing a postulated but not necessarily true likelihood function is denoted as *pseudo-maximum-likelihood* (PML) or a *quasi-maximum-likelihood* estimator. The relationship to method of moments shows that the PML estimator is consistent and asymptotically normal under the weaker conditions for GMM.

Consider a log-likelihood function given by

$$\log L(\theta) = \sum_{t=1}^{T} \log L_t(\theta \mid y_t),$$

where $L_t(\theta \mid y_t)$ is the likelihood contribution for observation t given the data. First order conditions for a maximum are given by the likelihood equations, $s(\theta) = \sum_{t=1}^{T} s_t(\theta) = 0$. Now note, that these equations can be seen as a set of K sample moment conditions

$$g_T(\theta) = T^{-1} \sum_{t=1}^{T} s_t(\theta) = T^{-1} \sum_{t=1}^{T} \frac{\partial \log L_t(\theta)}{\partial \theta} = 0, \tag{B2-1}$$

to which $\widehat{\theta}_{ML}$ is the unique MM solution. The population moment conditions corresponding to the sample moments in (B2-1) are given by

$$g(\theta_0) = E[s_t(\theta_0)] = 0, \tag{B2-2}$$

where $s_t(\theta) = f(y_t, \theta)$ in the GMM notation.

The MM estimator, $\widehat{\theta}_{MM}$, is the unique solution to (B2-1) and it is known to be a consistent estimator of θ_0 as long as the population moment conditions in (B2-2) are true. This implies that even if the likelihood function, $\log L(\theta)$, is misspecified, then the MM or PML estimator, $\widehat{\theta}_{MM} = \widehat{\theta}_{PML}$, is consistent as long as the moment conditions (B2-2) are satisfied. This shows that the ML estimator can be consistent even if the likelihood function is misspecified; we may say that the likelihood analysis shows some *robustness* to the specification.

It follows from the properties of GMM that the asymptotic variance of the PML estimator is not the inverse information, but is given by $V_{PML} = (D'S^{-1}D)^{-1}$ from (20). Under correct specification of the likelihood function this expression can be shown to simplify to $V_{ML} = S = I(\theta)^{-1}$. In a given application where we think that the likelihood function is potentially misspecified, it may be a good idea to base inference on the larger PML variance, V_{PML}, rather the ML variance, V_{ML}, and it is often a good idea to compare the two variances. A big difference may suggest that the likelihood function is misspecified.

Hypothesis testing on $\widehat{\theta}_{GMM}$ can be based on the asymptotic distribution:

$$\widehat{\theta}_{GMM} \overset{a}{\sim} N(\theta_0, T^{-1}\widehat{V}).$$

An estimator of the asymptotic variance is given by $\widehat{V} = \left(D_T' S_T^{-1} D_T\right)^{-1}$, where D_T is the sample average of the first derivatives in (19) and S_T is an estimator of $S = T \cdot V[g_T(\theta)]$. If the observations are independent, a consistent estimator is

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta) f(w_t, z_t, \theta)', \tag{21}$$

see the discussion of weight matrix estimation in Box 3.

2.4.1 Test of Overidentifying Moment Conditions

Recall that K moment conditions were sufficient to obtain a MM estimator of the K parameters in θ. If the estimation is based on $R > K$ moment conditions, we can test the validity of the $R - K$ overidentifying moment conditions. The intuition is that by MM estimation we can set K moment conditions equal to zero, but if all R moment conditions are valid then the remaining $R - K$ moments should also be close to zero. If a sample moment condition is far from zero it indicates that it is violated by the data.

It follows from (16) that

$$g_T(\theta_0) \overset{a}{\sim} N(0, T^{-1}S).$$

If we use the optimal weights, $W_T^{opt} \to S^{-1}$, then $\widehat{\theta}_{GMM} \to \theta_0$, and

$$\xi_J = T \cdot g_T(\widehat{\theta}_{GMM})' W_T^{opt} g_T(\widehat{\theta}_{GMM}) = T \cdot Q_T(\theta) \to \chi^2(R-K).$$

This is the standard result that the square of a normal variable is χ^2. The intuitive reason for the $R - K$ degrees of freedom (and not R, which is the dimension of $g_T(\theta)$) is that we have used K parameters to minimize $Q_T(\theta)$. If we wanted we could put K moment conditions equal to zero, and they would not contribute to the test.

The test is known as the *J-test* or the *Hansen test for overidentifying restrictions.* In linear models, to which we return below, the test is often referred to as the *Sargan test.* It is important to note that ξ_J does not test the validity of model *per se*; and in particular it is not a test of whether the underlying economic theory is correct. The test considers whether the $R - K$ overidentifying conditions are correct, given identification using K moments.

We cannot see from ξ_J directly, which moments that causes the rejection. If the test rejects, however, we may try to remove some moment conditions, reestimate and reconsider the statistic. This would give some indication of the problematic moment conditions or problematic instruments.

It can be shown that if the second estimation is based on $R_1 < R$ moment conditions, with $R_1 \geq K$ for identification, then it holds that the difference in the $J-$statistic is distributed as a $\chi^2(R - R_1)$ under the null of R valid moment conditions. In practice it is important to base both estimations on the same estimate of S, i.e. that the second estimation reuses R_1 rows and columns of S_T from the first estimation. These incremental $J-$statistics are sometimes referred to as $C-$tests[1].

2.5 Computational Issues

To obtain the efficient GMM estimator we need an optimal weight matrix. But note from (21) that the weight matrix depends on the parameters in general, and to estimate the optimal weight matrix we need a consistent estimator of θ_0.

First-step GMM estimator: Initially, we could choose some weight matrix, e.g. an identity $W_{[1]} = I_R$, and find the *first-step GMM estimator*

$$\widehat{\theta}_{[1]} = \arg\min_{\theta} g_T(\theta)' W_{[1]} g_T(\theta).$$

This estimator is consisten but inefficient.

Two-step efficient GMM estimator: To obtain an efficient estimator we could estimate the optimal weight matrix, $W_{[2]}^{opt}$, based on the consistent first step estimator, $\widehat{\theta}_{[1]}$. And given the optimal weight matrix we can find the efficient GMM estimator

$$\widehat{\theta}_{[2]} = \arg\min_{\theta} g_T(\theta)' W_{[2]}^{opt} g_T(\theta).$$

This procedure is denoted *two-step efficient GMM*. This estimator is not unique as it depends on the choice of the initial weight matrix $W_{[1]}$.

Iterated GMM estimator: Looking at the two-step procedure, it is natural to make another iteration. That is to reestimate the optimal weight matrix, $W_{[3]}^{opt}$, based on $\widehat{\theta}_{[2]}$, and then update the optimal estimator $\widehat{\theta}_{[3]}$. If we switch between estimating $W_{[\cdot]}^{opt}$ and $\widehat{\theta}_{[\cdot]}$ until convergence (i.e. that the parameters do not change from one iteration to the next) we obtain the so-called *iterated GMM estimator*, which does not depend on the initial weight matrix, $W_{[1]}$.

The two approaches are asymptotically equivalent. The intuition is that the estimators of θ and W^{opt} are consistent, so for $T \to \infty$ the iterated GMM estimator will converge in two iterations. For a given data set, however, there may be gains from the iterative procedure.

Continuously updated GMM estimator: A third approach is to recognize from the outset that the weight matrix depends on the parameters, and to reformulate the GMM criteria as

$$Q_T(\theta) = g_T(\theta)' W_T(\theta) g_T(\theta),$$

[1] The $C-$test gives a way of implementing misspecification tests. If we have $E(\epsilon_t \mid x_t) = 0$ for identification, homoskedasticity would imply $E(\epsilon_t^2 \mid x_t)$ being constant, which could be formulated as a moment condition.

Box 3: HC and HAC Weight Matrix Estimation

The optimal weight matrix is given by $W_T^{opt} = S_T^{-1}$, where S_T is a consistent estimator of

$$S = T \cdot V[g_T(\theta)] = T \cdot V\left[\frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta)\right] = \frac{1}{T} \cdot V\left[\sum_{t=1}^{T} f(w_t, z_t, \theta)\right]. \qquad \text{(B3-1)}$$

How to construct this estimator depends on the properties of the data. If the data are independent, then the variance of the sum is the sum of the variances, and we get that

$$S = \frac{1}{T}\sum_{t=1}^{T} V\left[f(w_t, z_t, \theta)\right] = \frac{1}{T}\sum_{t=1}^{T} E\left[f(w_t, z_t, \theta) f(w_t, z_t, \theta)'\right].$$

A natural estimator is

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta) f(w_t, z_t, \theta)'. \qquad \text{(B3-2)}$$

This is robust to heteroskedasticity by construction and is often referred to as the *heteroskedasticity consistent* (HC) variance estimator.

In the case of autocorrelation, $f(w_t, z_t, \theta)$ and $f(w_s, z_s, \theta)$ are correlated, and the variance of the sum in (B3-1) is not the sum of variances but includes contributions from all the covariances

$$S = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[f(w_t, z_t, \theta) f(w_s, z_s, \theta)'\right].$$

This is the so-called long-run variance of $f(\cdot)$ and the estimators are referred to as the class of *heteroskedasticity and autocorrelation consistent* (HAC) variance estimators. To describe the HAC estimators, first define the $R \times R$ sample covariance matrix between $f(w_t, z_t, \theta)$ and $f(w_{t-j}, z_{t-j}, \theta)$,

$$\Gamma_T(j) = \frac{1}{T}\sum_{t=j+1}^{T} f(w_t, z_t, \theta) f(w_{t-j}, z_{t-j}, \theta)'.$$

The natural estimator of S is then given by

$$S_T = \sum_{j=-T+1}^{T-1} \Gamma_T(j) = \Gamma_T(0) + \sum_{j=1}^{T-1} \left(\Gamma_T(j) + \Gamma_T(j)'\right), \qquad \text{(B3-3)}$$

where $\Gamma_T(0)$ is the HC estimator in (B3-2), and the last equality follows from the symmetry of the autocovariances, $\Gamma_T(j) = \Gamma_T(-j)'$.

Note, however, that we cannot consistently estimate as many covariances as we have observations and the simple estimator in (B3-3) is not necessarily positive definite. The trick is to put a weight w_j on autocovariance j, and to let the weights go to zero as j increases. This class of so-called *kernel estimators* can be written as

$$S_T = \Gamma_T(0) + \sum_{j=1}^{T-1} w_j \left(\Gamma_T(j) + \Gamma_T(j)'\right), \qquad \text{(B3-4)}$$

where $w_j = k\left(\frac{j}{B}\right)$. The function $k(\cdot)$ is a chosen kernel function and the constant B is referred to as a bandwidth parameter. A simple choice is the *Bartlett kernel*, where

$$w_j = k\left(\frac{j}{B}\right) = \begin{cases} 1 - \frac{j}{B} & \text{for} \quad \frac{j}{B} > 0 \\ 0 & \text{for} \quad \frac{j}{B} \leq 0 \end{cases}.$$

Box 4: box 3 continued

For this kernel, the weights decrease linearly with j and the weights are zero for $j \geq B$. We can think of the bandwidth parameter B as the maximum order of autocorrelation taken into account by the estimator. This estimator is also known as the *Newey-West estimator*. Other kernel functions exist which let the weights go to zero following some smooth pattern.

For a given kernel the bandwidth has to be chosen. If the maximum order of autocorrelation is unknown, then the (asymptotically optimal) bandwidth can be estimated from the data in an automated procedure; this is implemented in many software programs.

Finally, note that the HAC covariance estimator can also be used for calculating the standard errors for OLS estimates. This makes inference robust to autocorrelation.

and minimize this with respect to θ. This procedure, which is called the *continuously updated GMM estimator*, is never possible to solve analytically, but it can easily be implemented on a computer using numerical optimization.

3 Instrumental Variables Estimation

In many applications, the function in the moment condition has the specific form,

$$f(w_t, z_t, \theta) = z_t \cdot u(w_t, \theta),$$

where an $R \times 1$ vector of instruments, z_t, is multiplied by the 1×1 so-called *disturbance term*, $u(w_t, \theta)$. We could think of $u(w_t, \theta)$ as being the GMM equivalent of an error term, and the condition

$$g(\theta_0) = E[z_t \cdot u(w_t, \theta)] = 0, \tag{22}$$

states that the instruments should be uncorrelated with the disturbance term of the model. The class of estimators derived from (22) is referred to as *instrumental variables estimators.*

In some cases $u(w_t, \theta_0)$ is a non-linear function in which case the numerical procedures outlined above can be applied. In some cases, however, $u(w_t, \theta_0)$ is a linear function and the calculations may be done analytically. Below we illustrate this for the linear instrumental variables estimation

3.1 Linear Instrumental Variables Estimation

In this section we go through some of the details of GMM estimation for a linear regression model. The simplest case of the OLS estimator was considered in Example 2. Here we begin by restating the case for an exactly identified IV estimator also considered in Example 4; we then extend to overidentified cases.

3.1.1 Exact Identification

Consider again the case considered in Example 3, i.e. a partitioned regression

$$y_t = x_{1t}'\gamma_0 + x_{2t}'\delta_0 + \epsilon_t, \quad t = 1, 2, ..., T,$$

where

$$\begin{aligned} E[x_{1t}\epsilon_t] &= 0 \quad (K_1 \times 1) & (23) \\ E[x_{2t}\epsilon_t] &\neq 0 \quad (K_2 \times 1). & (24) \end{aligned}$$

The K_1 variables in x_{1t} are *predetermined*, while the $K_2 = K - K_1$ variables in x_{2t} are *endogenous.*

To obtain identification of the parameters we assume that there exists K_2 new variables, z_{2t}, that are correlated with x_{2t} but uncorrelated with the errors:

$$E[z_{2t}\epsilon_t] = 0. \tag{25}$$

Using the notation

$$\underset{(K\times 1)}{x_t} = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}, \quad \underset{(K\times 1)}{z_t} = \begin{pmatrix} x_{1t} \\ z_{2t} \end{pmatrix} \quad \text{and} \quad \beta_0 = \begin{pmatrix} \gamma_0 \\ \delta_0 \end{pmatrix},$$

we have K moment conditions:

$$g(\beta_0) = E[z_t u_t] = E[z_t \epsilon_t] = E[z_t \left(y_t - x_t'\beta_0\right)] = 0,$$

where $u(y_t, x_t, \beta_0) = y_t - x_t'\beta_0$ is the error term from the linear regression model.

We can write the corresponding sample moment conditions as

$$g_T(\widehat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} z_t \left(y_t - x_t'\widehat{\beta}\right) = \frac{1}{T}Z'\left(Y - X\widehat{\beta}\right) = 0, \tag{26}$$

where capital letters denote the usual stacked matrices

$$\underset{(T\times 1)}{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, \quad \underset{(T\times K)}{X} = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{pmatrix}, \quad \text{and} \quad \underset{(T\times K)}{Z} = \begin{pmatrix} z_1' \\ z_2' \\ \vdots \\ z_T' \end{pmatrix}.$$

The MM estimator is the unique solution:

$$\widehat{\beta}_{MM} = \left(\sum_{t=1}^{T} z_t x_t'\right)^{-1} \sum_{t=1}^{T} z_t y_t = \left(Z'X\right)^{-1} Z'Y,$$

provided that the $K \times K$ matrix $Z'X$ can be inverted. We note that if the number of new instruments equals the number of endogenous variables, then the GMM estimator coincides with the simple IV estimator.

3.1.2 Overidentification

Now assume that we want to introduce more instruments and let $z_t = (x_1', z_2')'$ be an $R \times 1$ vector with $R > K$. In this case $Z'X$ is no longer invertible and the MM estimator does not exist. Now we have R moments

$$g_T(\beta) = \frac{1}{T} \sum_{t=1}^{T} z_t \left(y_t - x_t'\beta \right) = \frac{1}{T} Z' \left(Y - X\beta \right),$$

and we cannot solve $g_T(\beta) = 0$ directly. Instead, we want to derive the GMM estimator by minimizing the criteria function

$$\begin{aligned} Q_T(\beta) &= g_T(\beta)' W_T g_T(\beta) \\ &= \left(T^{-1} Z' \left(Y - X\beta \right) \right)' W_T \left(T^{-1} Z' \left(Y - X\beta \right) \right) \\ &= T^{-2} \left(Y'ZW_TZ'Y - 2\beta'X'ZW_TZ'Y + \beta'X'ZW_TZ'X\beta \right), \end{aligned}$$

for some weight matrix W_T. We take the first derivative, and the GMM estimator is the solution to the K equations

$$\frac{\partial Q_T(\beta)}{\partial \beta} = -2T^{-2}X'ZW_TZ'Y + 2T^{-2}X'ZW_TZ'X\beta = 0$$

that is

$$\widehat{\beta}_{GMM}(W_T) = \left(X'ZW_TZ'X \right)^{-1} X'ZW_TZ'Y.$$

The estimator depends on the weight matrix, W_T. To estimate the optimal weight matrix, $W_T^{opt} = S_T^{-1}$, we use the estimator in (21), that is

$$S_T = \frac{1}{T} \cdot \sum_{t=1}^{T} f(w_t, z_t, \theta) f(w_t, z_t, \theta)' = \frac{1}{T} \sum_{t=1}^{T} \widehat{\epsilon}_t^2 z_t z_t', \tag{27}$$

which allows for general heteroskedasticity of the disturbance term. The efficient GMM estimator is given by

$$\widehat{\beta}_{GMM} = \left(X'ZS_T^{-1}Z'X \right)^{-1} X'ZS_T^{-1}Z'Y,$$

where we note that any scale factor in the weight matrix, e.g. T^{-1}, cancels.

For the asymptotic distributions, we recall that

$$\widehat{\beta}_{GMM} \overset{a}{\sim} N \left(\beta_0, T^{-1} \left(D'S^{-1}D \right)^{-1} \right).$$

The derivative is given by

$$\underset{(R \times K)}{D_T} = \frac{\partial g_T(\beta)}{\partial \beta'} = \frac{\partial \left(T^{-1} \sum_{t=1}^{T} z_t \left(y_t - x_t'\beta \right) \right)}{\partial \beta'} = -T^{-1} \sum_{t=1}^{T} z_t x_t',$$

so the variance of the estimator becomes

$$\begin{aligned} V\left[\widehat{\beta}_{GMM}\right] &= T^{-1}\left(D_T' W_T^{opt} D_T\right)^{-1} \\ &= T^{-1}\left(\left(-T^{-1}\sum_{t=1}^{T} x_t z_t'\right)\left(T^{-1}\sum_{t=1}^{T}\widehat{\epsilon}_t^2 z_t z_t'\right)^{-1}\left(-T^{-1}\sum_{t=1}^{T} z_t x_t'\right)\right)^{-1} \\ &= \left(\sum_{t=1}^{T} x_t z_t'\right)^{-1}\sum_{t=1}^{T}\widehat{\epsilon}_t^2 z_t z_t'\left(\sum_{t=1}^{T} z_t x_t'\right)^{-1}. \end{aligned}$$

We recognize this expression as the *heteroskedasticity consistent* (HC) variance estimator of White. Using GMM with the allowance for heteroskedastic errors will thus automatically produce heteroskedasticity consistent standard errors.

If we assume that the error terms are IID, then the optimal weight matrix in (27) simplifies to

$$S_T = \frac{\widehat{\sigma}^2}{T}\sum_{t=1}^{T} z_t z_t' = T^{-1}\widehat{\sigma}^2 Z'Z, \tag{28}$$

where $\widehat{\sigma}^2$ is a consistent estimator for σ^2. In this case the efficient GMM estimator becomes

$$\begin{aligned} \widehat{\beta}_{GMM} &= \left(X'ZS_T^{-1}Z'X\right)^{-1} X'ZS_T^{-1}Z'Y. \\ &= \left(X'Z\left(T^{-1}\widehat{\sigma}^2 Z'Z\right)^{-1} Z'X\right)^{-1} X'Z\left(T^{-1}\widehat{\sigma}^2 Z'Z\right)^{-1} Z'Y \\ &= \left(X'Z\left(Z'Z\right)^{-1} Z'X\right)^{-1} X'Z\left(Z'Z\right)^{-1} Z'Y. \end{aligned}$$

Notice that the efficient GMM estimator is identical to the generalized IV estimator and the two stage least squares (2SLS) estimator. This shows that the 2SLS estimator is the efficient GMM estimator if the error terms are IID. The variance of the estimator is

$$V\left[\widehat{\beta}_{GMM}\right] = T^{-1}\left(D_T' S_T^{-1} D_T\right)^{-1} = \widehat{\sigma}^2 (X'Z\left(Z'Z\right)^{-1} Z'X)^{-1},$$

which again coincides with the 2SLS variance.

4 Empirical Examples

In this section we present a number of empirical illustrations.

4.1 Software Installation

To illustrate the GMM methodology, I have written as a small GMM module for OxMetrics. Due to the generality and the non-linearity of GMM, estimation always require some degree of programming; and the practitioner has to make decisions on details in the implementation. The OxMetrics module has similarities with the PcGive module for ML estimation although it is less advanced.

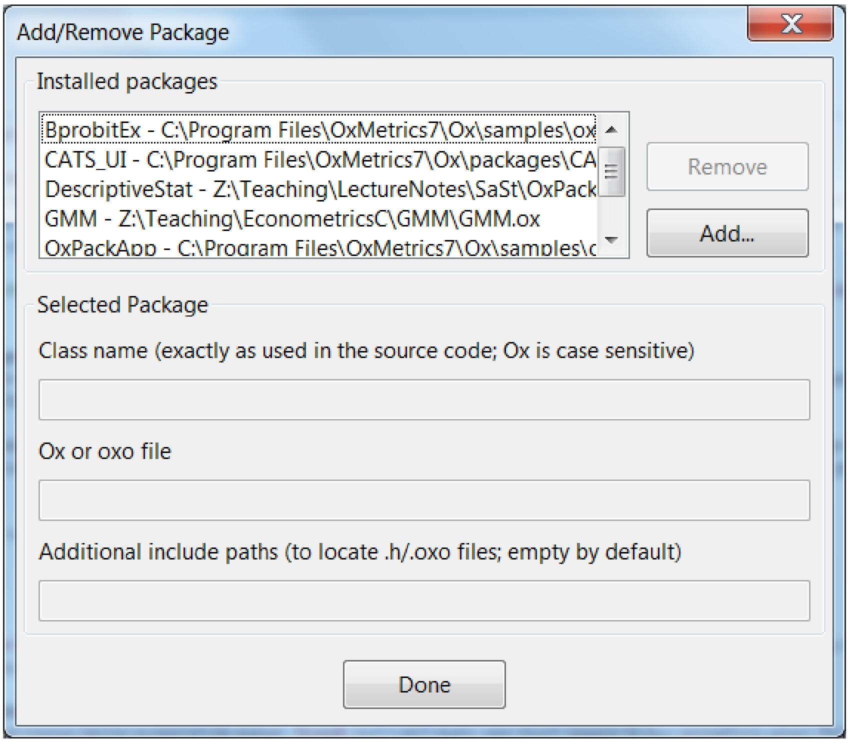

Figure 1: *Installation of the GMM module via OxPack.*

To install the program unpack the files in `GMM_3.x.ZIP` to a folder in which you are allowed to write, e.g.

$$\texttt{C:\textbackslash Econometrics\textbackslash GMM\textbackslash} \tag{29}$$

or similar.

Inside OxMetrics choose the OxPack module and select the `[Add/Remove Package...]` from the `[Package]` menu. Now choose `[Add]` and find the file `GMM.ox` you have just installed and add it. Press `[Done]` to close the window, see Figure 1

The GMM module should now be running and you should see the message `GMM 3.x session started...` in the results window.

4.2 Simple Estimators

To illustrate the module we first consider some simple examples. We use a data set where DC denote the change in the log of Danish consumption and DY denote the change in the log of disposable income, for the sample $1971:2-2003:2$.

4.2.1 OLS Estimation

To implement the OLS regression

$$\mathsf{DC}_t = \beta_1 + \beta_2 \mathsf{DY}_t + \epsilon_t,$$

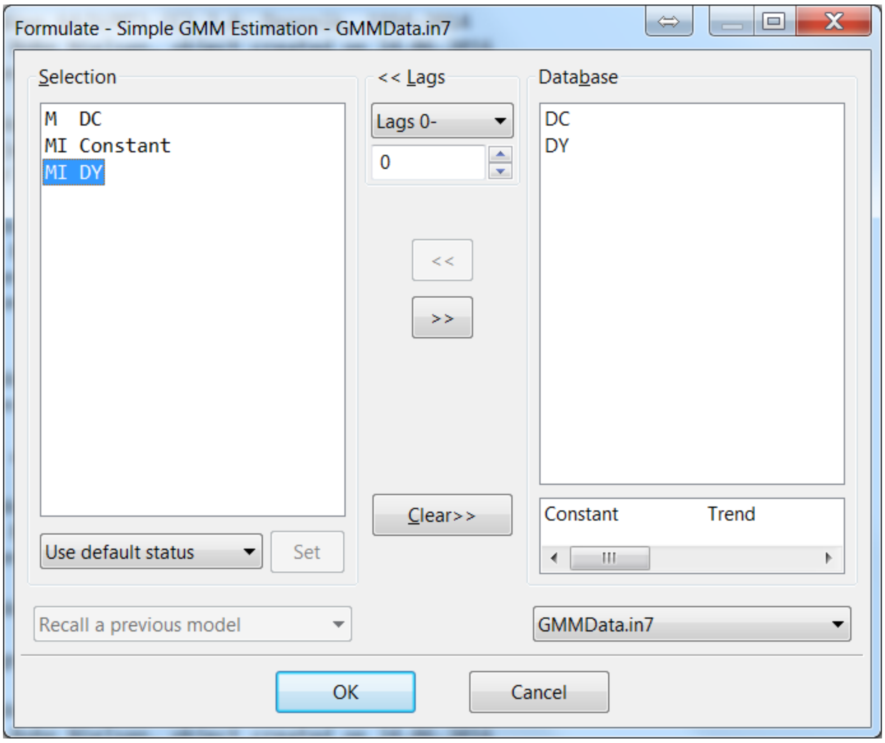

Figure 2: *Selecting variables for the GMM estimation.*

as a GMM estimator, we note that $Y_t = (\mathsf{DC}_t, 1, \mathsf{DY}_t)'$ are *model variables*, $Z_t = (1, \mathsf{DY}_t)'$ are *instruments*, and the moment conditions are given by

$$E[\epsilon_t Z_t] = 0.$$

In the GMM module we first choose the variables as in PcGive and set the relevant variables as model variables and instruments, see Figure 2.

We then get a window with options for the estimation, see Figure 3, and for the current example we choose Iterated GMM. The choice of weight matrix is not important for the exactly identified estimators, but they will affect the estimated variances. To reproduce the OLS results we choose IID weight matrix.

In the next window, we choose the largest possible sample 1971(2)-2003(2). To specify the moment conditions we have to do a little programming. For the programming, the model variables are renamed as the $T \times 1$ column vectors Y[0], Y[1], and Y[2]; the instruments are renamed as the $T \times 1$ vectors Z[0] and Z[1]; while the parameters are denoted {beta_1}, {beta_2} etc. Parameters may be renamed, as long as the names are enclosed in $\{\cdot\}$. The default in the programming window is the OLS-type moment conditions, see Figure 4.

The first line defines the $T \times 1$ vector of residuals,

$$\mathsf{Uvec} = \mathsf{Y}[0] - \{\mathsf{beta\_1}\} * \mathsf{Y}[1] - \{\mathsf{beta\_2}\} * \mathsf{Y}[2];$$

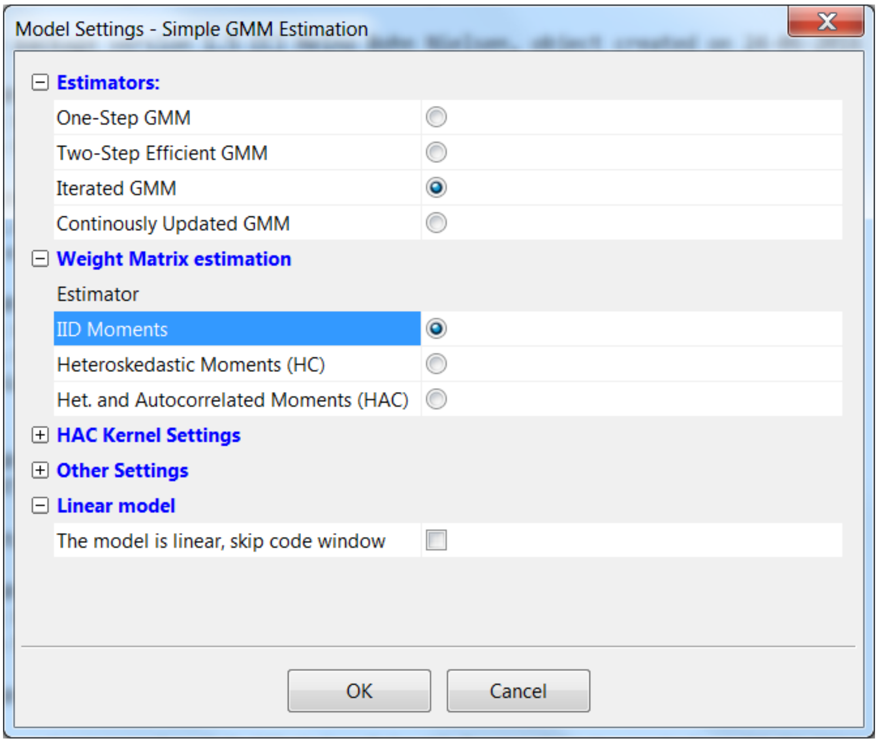

Figure 3: *Options for GMM estimation.*

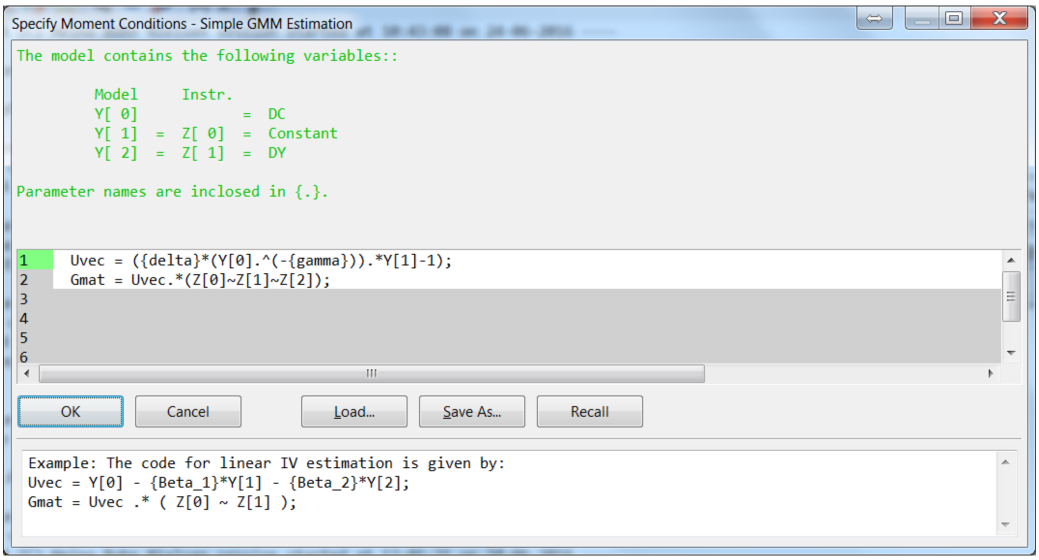

Figure 4: *Programming the GMM moment conditions.*

	Estimator	Weight	β_0	β_1	T	ξ_J	DF	$p-val$
OLS	Iterated GMM	IID	0.0024 (0.0016)	0.2100 (0.0611)	129	0.000	–	–
OLS	Iterated GMM	HC	0.0024 (0.0015)	0.2100 (0.0921)	129	0.000	–	–
OLS	Iterated GMM	HAC	0.0024 (0.0011)	0.2100 (0.1086)	129	0.000	–	–
IV	Iterated GMM	IID	0.0037 (0.0019)	−0.1340 (0.1957)	127	9.081	3	0.028
IV	Iterated GMM	HC	0.0049 (0.0018)	−0.1761 (0.1795)	127	6.018	3	0.111
IV	Iterated GMM	HAC	0.0039 (0.0011)	−0.1715 (0.1149)	127	2.581	3	0.462

Table 1: *GMM estimation of simple models. Standard errors in parentheses. 'IID' denotes independent and identically distributed moments. 'HC' denotes the estimator allowing for heteroskedasticity of the moments. 'HAC' denotes the estimator allowing for heteroskedasticity and autocorrelation. In the implementatioon of the HAC estimator we allow for autocorrelation of order 12 using the Bartlett kernel. 'DF' is the number of overidentifying moments for the Hansen test, ξ_J, and 'p-val' is the corresponding p-value.*

where element t is just the residual $\epsilon_t = \mathsf{DC}_t - \beta_1 - \beta_2 \mathsf{DY}_t$. The second line defines the $T \times 2$ matrix of moments

$$\mathsf{Gmat} = \mathsf{Uvec}. * (\mathsf{Z[0]}\tilde{}\mathsf{Z[1]});$$

in which row t is given by $f_t' = \epsilon_t \cdot (1, \mathsf{DY}_t)$. The notation $.*$ is Ox code for element by element multiplication rather than matrix multiplication, and ˜ is concatenation of column vectors. The reported results for the 4 estimators are identical, and the iterated GMM results are reported in first row of Table 1. We note that the criteria function is zero because the model is just identified.

To make the variances robust to heteroskedasticity, we just redo the analysis with the HC weight matrix. We note that the estimators are the same but the $t-$values are slightly different. We can also make the inference robust to autocorrelation by choosing the HAC weight matrix, which again changes the variances. Here we use a Bartlett kernel with $B = 12$ lags, so that the weights in (B3-4) are given by $w_1 = 11/12 = 0.917, w_2 = 10/12 = 0.833, ..., w_{11} = 1/12 = 0.083$, while $w_j = 0$ for $j \geq 12$. These GMM-type corrections of the OLS variance are standard in econometric software packages and they are also available in PcGive.

4.2.2 Two-Stage Least Squares

To implement instrumental variables estimation we assume that DY_t is endogenous and we instrument it with lags, $\mathsf{DY}_{t-1}, \mathsf{DY}_{t-2}, \mathsf{DC}_{t-1}, \mathsf{DC}_{t-2}$. In this case we change the list of instruments to $Z_t = (1, \mathsf{DY}_{t-1}, \mathsf{DY}_{t-2}, \mathsf{DC}_{t-1}, \mathsf{DC}_{t-2})'$, and the moment conditions to

$$\mathsf{Gmat} = \mathsf{Uvec.} * (\mathsf{Z[0]}\tilde{}\,\mathsf{Z[1]}\tilde{}\,\mathsf{Z[2]}\tilde{}\,\mathsf{Z[3]}\tilde{}\,\mathsf{Z[4]});$$

The estimation is now over-identified with 3 over-identifying moment conditions. The optimal GMM estimator in the case of IID moments is Two-Stage Least Squares, reported in row 4 of Table 1. The reason that the coefficients change so much is either that DY_t is strongly endogenous and OLS is invalid, or that the instruments are weak. Looking at the first step estimates from the two-stages least squares estimation, we get

$$\mathsf{DY}_t = \underset{(2.08)}{0.0047} - \underset{(-3.40)}{0.3212}\mathsf{DY}_{t-1} - \underset{(-0.578)}{0.0538}\,\mathsf{DY}_{t-2} + \underset{(1.58)}{0.2010}\mathsf{DC}_{t-1} - \underset{(-1.20)}{0.1549}\mathsf{DC}_{t-2}.$$

The coefficient of determination is $R^2 = 0.12$, and the Wald F-test for all coefficients equal to zero is 4.345 and produces a $p-$value of 0.003. This indicates that the identification is formally valid, but the instruments are probable relatively weak.

The weight matrix now play a role for the optimal GMM estimators, and changing the weight matrix produces different estimators, see Table 1.

4.3 Optimal Monetary Policy

Now consider a more interesting application where IV and GMM are relevant. Many authors have suggested that monetary policy can be described by a reaction function in which the policy interest rate reacts on the deviation of expected future inflation from a constant target value, and the output gap, i.e. the deviation of real activity from potential. Let π_t denote the current inflation rate from the year before, and let π^* denote the constant inflation target of the central bank. Furthermore, let $\widetilde{y}_t = y_t - y_t^*$ denote a measure of the output gap. The reaction function for the policy rate r_t can then be written in a simple so-called Taylor-rule:

$$r_t = \alpha_0 + \alpha_1 \cdot E[\pi_{t+12} - \pi^* \mid \mathcal{I}_t] + \alpha_2 \cdot E[\widetilde{y}_t \mid \mathcal{I}_t], \tag{30}$$

where α_0 is interpretable as the target value of r_t in equilibrium. We have assumed that the relevant forecast horizon of the central bank is 12 month, and $E[\pi_{t+12} \mid \mathcal{I}_t]$ is the best forecast of inflation one year ahead given the information set of the central bank, $\mathcal{I}_t$. The forecast horizon should reflect the lag of the monetary transmission. The parameter α_1 is central in characterizing the behavior of the central bank. If $\alpha_1 > 1$ then the central bank will increase the real interest rate to stabilize inflation, while the a reaction $\alpha_1 \leq 1$ is formally inconsistent with inflation stabilization.

The relevant central bank forecasts cannot be observed, and inserting observed values we obtain the model

$$r_t = a_0^* + \alpha_1 \cdot \pi_{t+12} + \alpha_2 \cdot \widetilde{y}_t + u_t, \tag{31}$$

where the constant term $\alpha_0^* = (\alpha_0 - \alpha_1 \pi^*)$ now includes the constant inflation target, π^*. Also note that the new error term contains the forecast errors:

$$u_t = \alpha_1 \cdot (E[\pi_{t+12} \mid \mathcal{I}_t] - \pi_{t+12}) + \alpha_2 \cdot (E[\widetilde{y}_t \mid \mathcal{I}_t] - \widetilde{y}_t)\,. \tag{32}$$

The model in (31) is a linear model in (ex post) observed quantities, π_{t+12} and $\widetilde{y}_t$, but we cannot apply simple linear regression because the error term u_t is correlated with the explanatory variables. If we assume that the forecasts are rational, however, then all variables in the information set of the central bank at time t should be uninformative on the forecasts errors, and

$$E[u_t \mid \mathcal{I}_t] = 0.$$

This zero conditional expectation implies the unconditional moment conditions

$$E[u_t z_t] = 0, \tag{33}$$

for all variables $z_t \in \mathcal{I}_t$ included in the formation set, and we can estimate the parameters in (30) by linear instrumental variables estimation. Using the model formulation, the moment conditions have the form

$$E[u_t z_t] = E\left[\{r_t - \alpha_0^* - \alpha_1 \cdot \pi_{t+12} + \alpha_2 \cdot \widetilde{y}_t\} \cdot z_t\right] = 0,$$

for instruments $z_{1t}, ..., z_{Rt}$. We need at least $R = 3$ instruments to estimate the three parameters $\theta = (\alpha_0^*, \alpha_1, \alpha_2)'$. As instruments we should choose variables that can explain the forecasts $E[\pi_{t+12} \mid \mathcal{I}_t]$ and $E[\widetilde{y}_t \mid \mathcal{I}_t]$ while at the same time being uncorrelated with the disturbance term, u_t. Put differently, we could choose variables that the central bank use in their forecasts, but which they do not react directly upon. As an example the long-term interest rate is a potential instrument if it is informative on future inflation–but if the central bank reacts directly on the movements of the bond rate, then an orthogonality condition in (33) is violated and the bond rate should have been included in the reaction function. In a time series model lagged variables are always possible instruments, but in many cases they are relatively weak and they often have to be augmented with other variables.

To illustrate estimation, we consider a data set for US monetary policy under Greenspan, with effective sample $1988:1-2005:8$. We use the (average effective) Federal funds rate, ff_t, to measure the policy interest rate, r_t, and the CPI inflation rate year-over-year, inf_t, to measure π_t. As a measure of the output gap, $\widetilde{y}_t = y_t - y_t^*$, we use the deviation of capacity utilization from the average, capgap_t, so that large values imply high activity; and we expect $\alpha_2 > 0$. The time series are illustrated in Figure 6. For most of the period the Federal funds rate in (A) seems to be positively related to the capacity utilization in (C). For some periods the effect from inflation is also visible–e.g. around the year 2000 where the temporary interest rate increase seems to be explained by movements in inflation.

To estimate the parameters we choose a set of instruments consisting of a constant term and lagged values of the interest rate, inflation, and capacity utilization. For the presented results we use lag $1-6$ plus lag 9 and 12 of all variables:

$$z_t = (1, r_{t-1}, ..., r_{t-6}, r_{t-9}, r_{t-12}, \pi_{t-1}, ..., \pi_{t-6}, \pi_{t-9}, \pi_{t-12}, \widetilde{y}_{t-1}, ..., \widetilde{y}_{t-6}, \widetilde{y}_{t-9}, \widetilde{y}_{t-12})'.$$

That gives a total of $R = 25$ moment conditions to estimate the 3 parameters. The formulation window is given in Figure 5

If we assume that the moments are IID, then we can estimate the optimal weight matrix by (28) and the GMM estimator simplifies again to the two-stage least squares. The estimation results are presented in row (M1) in Table 2. We note that α_1 is significantly larger than one, indicating inflation stabilization, and there is a significant effect from the capacity utilization, $\alpha_2 > 0$. We have 22 overidentifying moment conditions and the Hansen test for overidentification of $\xi_J = 105$ is distributed as a $\chi^2(22)$ under correct specification. The statistic is much larger than the 5% critical value of 33.9 and we conclude that some of the moment conditions are violated. The values of the Federal funds rate predicted by the reaction function are illustrated in graph (D) together with the actual value Federal funds rate. We note that the observed interest rate is much more persistent than the prediction.

Allowing for heteroskedasticity of the moments produce the (iterated GMM) estimates reported in row (M2). These results are by and large identical to the results in row (M1).

The fact that u_t includes a 12-month forecast will automatically produce autocorrelation, and the optimal weight matrix should allow for autocorrelation up to lag 12. Using a HAC estimator of the weight matrix that allows autocorrelation of order 12 produces the results reported in row (M3). The parameter estimates are not too far from the previous models, although the estimate to inflation is a bit smaller. It is worth noting that the use of an autocorrelation consistent weight matrix makes the test for overidentification insignificant; and the 22 overidentifying conditions are overall accepted for this specification.

4.3.1 Interest Rate Smoothing.

The estimated Taylor rules based on (31) are unable to capture the high persistence of the actual Federal funds rate. In the literature many authors have suggested to reinterpret the Taylor rule as a target value and to model the actual reaction function as a partial adjustment process:

$$\begin{aligned} r_t^* &= \alpha_0^* + \alpha_1 \cdot E[\pi_{t+12} \mid \mathcal{I}_t] + \alpha_2 \cdot E[\widetilde{y}_t \mid \mathcal{I}_t] \\ r_t &= (1-\rho) \cdot r_t^* + \rho \cdot r_{t-1}. \end{aligned}$$

The two equations can be combined to produce

$$r_t = (1-\rho) \cdot \{\alpha_0^* + \alpha_1 \cdot E[\pi_{t+12} \mid \mathcal{I}_t] + \alpha_2 \cdot E[\widetilde{y}_t \mid \mathcal{I}_t]\} + \rho \cdot r_{t-1},$$

in which the actual interest rate depends on the lagged dependent variable. Replacing again expectations with actual observations we obtain an empirical model

$$r_t = (1-\rho) \cdot \{\alpha_0^* + \alpha_1 \cdot \pi_{t+12} + \alpha_2 \cdot \widetilde{y}_t\} + \rho \cdot r_{t-1} + u_t, \tag{34}$$

where the error term is given by (32) with α_i replaced by $\alpha_i(1-\rho)$ for $i = 1, 2$. The parameters in (34), $\theta = (\alpha_0^*, \alpha_1, \alpha_2, \rho)'$, can be estimated by linear GMM using the conditions in (33) with

$$u_t = r_t - (1-\rho) \cdot \{\alpha_0^* + \alpha_1 \cdot \pi_{t+12} + \alpha_2 \cdot \widetilde{y}_t\} - \rho \cdot r_{t-1}.$$

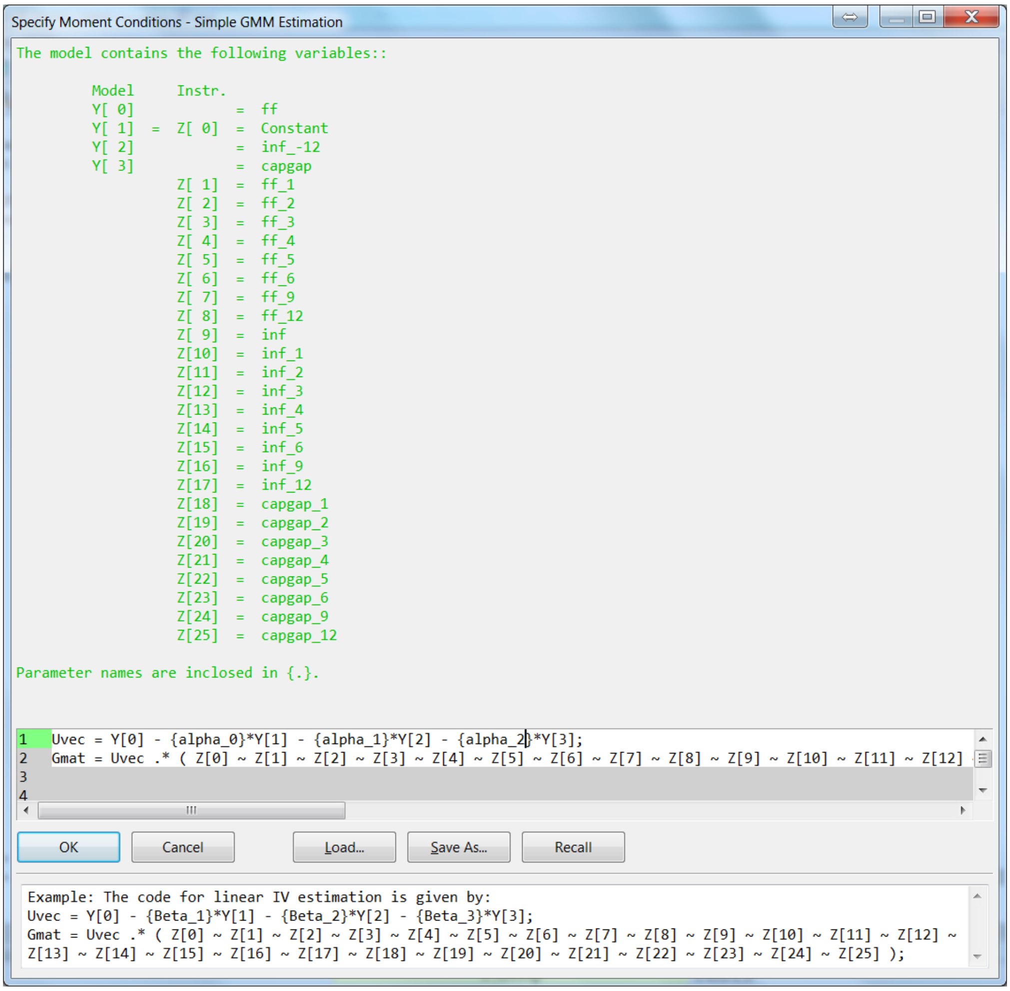

Figure 5: *Moment conditions for the forward looking monetary policy analysis.*

We note that the lagged Federal funds rate, r_{t-1}, is included in the information set at time t, so even if r_{t-1} is now a model variable it is a still included in the list of instruments. We say that it is instrument for itself. To estimate using the GMM module we reformulate the moment conditions as

$$\mathsf{Uvec} = \mathsf{Y}[0]-(1-\{\mathsf{rho}\})*(\{\mathsf{alpha\_0}\}*\mathsf{Y}[1]+\{\mathsf{alpha\_1}\}*\mathsf{Y}[3]+\{\mathsf{alpha\_2}\}*\mathsf{Y}[4])-\{\mathsf{rho}\}*\mathsf{Y}[2];$$

Rows (M4)−(M6) in Table 2 report the estimation results for the partial adjustment model (34). Allowing for interest rate smoothing changes the estimated parameters somewhat. We first note that the sensitivity to the business cycle, α_2, is markedly increased to values in the range $\frac{3}{4}$ to 1. The sensitivity to future inflation, α_1, depends more on the choice of weight matrix, ranging now from $1\frac{1}{4}$ to $1\frac{3}{4}$. We also note that the interest rate smoothing is very important. The coefficient to r_{t-1} is very close to one and the coefficient to the new information in r_t^* is below $\frac{1}{10}$. The predicted values are presented

		α_0^*	α_1	α_2	ρ	T	ξ_J	DF	$p-val$
(M1)	IID	0.5529 (0.4476)	1.4408 (0.1441)	0.3938 (0.0401)		212	105.576	22	0.000
(M2)	HC	0.4483 (0.3652)	1.5133 (0.1124)	0.3747 (0.0302)		212	54.110	22	0.000
(M3)	HAC	1.1959 (0.7506)	1.3333 (0.2143)	0.3551 (0.0620)		212	9.883	22	0.987
(M4)	IID	0.6483 (0.6469)	1.2905 (0.2093)	0.7108 (0.0732)	0.9213 (0.0102)	212	42.168	21	0.004
(M5)	HC	1.0957 (0.5923)	1.1881 (0.2052)	0.7254 (0.0561)	0.9240 (0.0094)	212	36.933	21	0.017
(M6)	HAC	−0.8355 (0.7314)	1.7385 (0.2459)	1.0714 (0.15584)	0.9284 (0.0108)	212	10.352	21	0.974

Table 2: *GMM estimation of monetary policy rules for the US. Standard errors in parentheses. 'IID' denotes independent and identically distributed moments. 'HC' denotes the estimator allowing for heteroskedasticity of the moments. 'HAC' denotes the estimator allowing for heteroskedasticity and autocorrelation. In the implementatioon of the HAC estimator we allow for autocorrelation of order 12 using the Bartlett kernel. 'DF' is the number of overidentifying moments for the Hansen test, ξ_J, and 'p-val' is the corresponding p-value.*

in graph (D), now capturing most of the persistence.

A coefficient to the lagged interest rate close to unity could reflect that the time series for r_t is very close to behaving as a unit root process. If this is the case then the tools presented here would not be valid, as Assumptions 1 and 2 would be violated. This case has not been seriously considered in the literature and is beyond the scope of this section.

4.4 The C-CAPM Model

To illustrate a non-linear GMM estimation we consider the (consumption based) capital asset pricing (C-CAPM) model of Hansen and Singleton (1982). A representative agent is assumed to choose an optimal consumption path, $c_t, c_{t+1}, ...$, by maximizing the present discounted value of lifetime utility, i.e.

$$\max \sum_{s=0}^{\infty} E\left[\delta^s \cdot u(c_{t+s}) \mid \mathcal{I}_t\right],$$

where $u(c_{t+s})$ is the utility of consumption, $0 \leq \delta \leq 1$ is a discount factor, and $\mathcal{I}_t$ is the information set at time t. The consumer can change the path of consumption relative to income by investing in a financial asset. Let A_t denote the financial wealth at the end of period t and let r_t be the implicit interest rate of the financial position. Then a feasible consumption path must obey the budget constraint

$$A_{t+1} = (1 + r_{t+1}) A_t + y_{t+1} - c_{t+1},$$

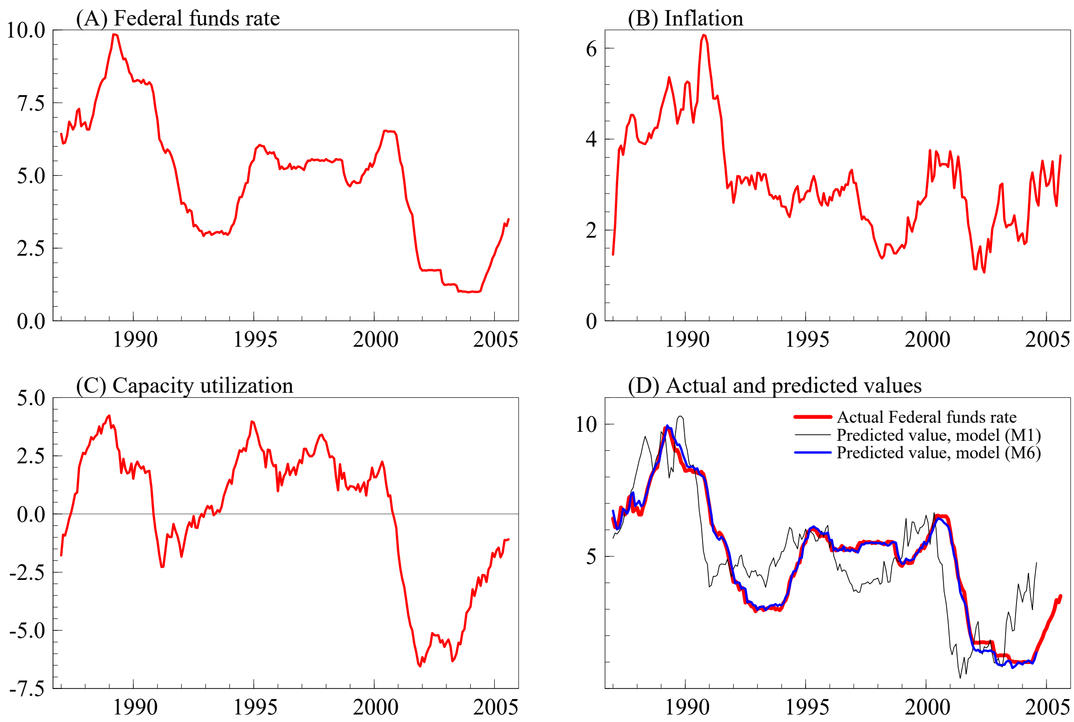

Figure 6: *Estimating reaction functions for US monetary policy for the Greenspan period.*

where y_t denotes labour income. The first order condition for this problem is given by

$$u'(c_t) = E\left[\delta \cdot u'(c_{t+1}) \cdot R_{t+1} \mid \mathcal{I}_t\right],$$

where $u'(\cdot)$ is the derivative of the utility function, and $R_{t+1} = 1 + r_{t+1}$ is the return factor.

To put more structure on the model, we assume a constant relative risk aversion (CRRA) utility function

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}, \quad \gamma < 1,$$

so that the first derivative is given by $u'(c_t) = c_t^{-\gamma}$. This formulation gives the explicit Euler equation:

$$c_t^{-\gamma} - E\left[\delta \cdot c_{t+1}^{-\gamma} \cdot R_{t+1} \mid \mathcal{I}_t\right] = 0$$

or alternatively

$$E\left[\delta \cdot \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \cdot R_{t+1} - 1 \mid \mathcal{I}_t\right] = 0. \tag{35}$$

The zero conditional expectation in (35) implies the unconditional moment conditions

$$E\left[f\left(c_{t+1}, c_t, R_{t+1}; z_t; \delta, \gamma\right)\right] = E\left[\left(\delta \cdot \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \cdot R_{t+1} - 1\right) z_t\right] = 0, \tag{36}$$

for all variables $z_t \in \mathcal{I}_t$ included in the formation set. The economic interpretation is that under rational expectations a variable in the information set must be uncorrelated to the expectation error.

We recognize (36) as a set of moment conditions of a non-linear instrumental variables model. Since we have two parameters to estimate, $\theta = (\delta, \gamma)'$, we need at least $R = 2$ instruments in z_t to identify θ. Note that the specification is fully theory driven, it is nonlinear, and it is not in a regression format. Moreover, the parameters we estimate are the "deep" parameters of the optimization problem.

To estimate the deep parameters, we have to choose a set of instruments z_t. Possible instruments could be variables from the joint history of the model variables, and here we take the 3×1 vector:

$$z_t = \left(1, \frac{c_t}{c_{t-1}}, R_t\right)'.$$

This choice would correspond to the three moment conditions

$$\begin{aligned}
E\left[\left(\delta \cdot \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \cdot R_{t+1} - 1\right)\right] &= 0 \\
E\left[\left(\delta \cdot \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \cdot R_{t+1} - 1\right)\left(\frac{c_t}{c_{t-1}}\right)\right] &= 0 \\
E\left[\left(\delta \cdot \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \cdot R_{t+1} - 1\right) R_t\right] &= 0,
\end{aligned}$$

for $t = 1, 2, ..., T$, but we could also extend with more lags.

To illustrate the procedures we use a data set similar to Hansen and Singleton (1982) consisting of monthly data for real consumption growth, c_t/c_{t-1}, and the real return on stocks, R_t, for the US $1959:3 - 1978:12$. We take $Y_t = (\frac{c_t}{c_{t-1}}, R_t)'$ as model variables and $Z_t = (1, \frac{c_{t-1}}{c_{t-2}}, R_{t-1})'$ as instruments and use a formulation with

$$\begin{aligned}
\mathsf{Uvec} &= (\{\mathsf{delta}\} * (\mathsf{Y}[0].\hat{}(-\{\mathsf{gamma}\})). * \mathsf{Y}[1] - 1); \\
\mathsf{Gmat} &= \mathsf{Uvec.} * (\mathsf{Z}[0]\tilde{}\mathsf{Z}[1]\tilde{}\mathsf{Z}[2]);
\end{aligned}$$

where .^ is the Ox function for element by element power.

Rows (N1)−(N3) in Table 3 report the estimation results for the nonlinear instrumental variable model where the weight matrix allows for heteroskedasticity of the moments. The models are estimated with, respectively, the two-step efficient GMM estimator, the iterated GMM estimator, and the continuously updated GMM estimator; and the results are by and large identical. The discount factor δ is estimated to be very close to unity, and the standard errors are relatively small. The coefficient of relative risk aversion, γ, on the other hand, is very poorly estimated, with very large standard errors. For the iterated GMM estimation in model (N2) the estimate is 1.0249 with a disappointing 95% confidence interval of $[-2.70; 4.75]$. We note that the Hansen test for the single overidentifying condition does not reject correct specification.

			Lags	δ	γ	T	ξ_J	DF	$p-val$
(N1)	2-Step	HC	1	0.9987 (0.0086)	0.8770 (3.6792)	237	0.434	1	0.510
(N2)	Iterated	HC	1	0.9982 (0.0044)	1.0249 (1.8614)	237	1.068	1	0.301
(N3)	CU	HC	1	0.9981 (0.0044)	0.9549 (1.8629)	237	1.067	1	0.302
(N4)	2-Step	HAC	1	0.9987 (0.0092)	0.8876 (4.0228)	237	0.429	1	0.513
(N5)	Iterated	HAC	1	0.9980 (0.0045)	0.8472 (1.8757)	237	1.091	1	0.296
(N6)	CU	HAC	1	0.9977 (0.0045)	0.7093 (1.8815)	237	1.086	1	0.297
(N7)	2-Step	HC	2	0.9975 (0.0066)	0.0149 (2.6415)	236	1.597	3	0.660
(N8)	Iterated	HC	2	0.9968 (0.0045)	−0.0210 (1.7925)	236	3.579	3	0.311
(N9)	CU	HC	2	0.9958 (0.0046)	−0.5526 (1.8267)	236	3.501	3	0.321
(N10)	2-Step	HAC	2	0.9970 (0.0068)	−0.1872 (2.7476)	236	1.672	3	0.643
(N11)	Iterated	HAC	2	0.9965 (0.0047)	−0.2443 (1.8571)	236	3.685	3	0.298
(N12)	CU	HAC	2	0.9952 (0.0048)	−0.9094 (1.9108)	236	3.591	3	0.309

Table 3: *Estimated Euler equations for the C-CAPM model. Standard errors in parentheses. '2-step' denotes the two-step efficient GMM estimator, where the initial weight matrix is a unit matrix. 'Iterated' denotes the iterated GMM estimator. 'CU' denotes the continously updated GMM estimator. 'Lags' is the number of lags in the instrument vector. 'DF' is the number of overidentifying moments for the Hansen test, ξ_J, and 'p-val' is the corresponding p-value.*

Rows (N4)−(N6) report estimation results for models where the weight matrix is robust to heteroskedasticity and autocorrelation. The results are basically unchanged.

We conclude that the used data set is not informative enough to empirically identify the coefficient of relative risk aversion, γ. One explanation could be that the economic model is in fact correct, but that we need stronger instruments to identify the parameter. One possible solution is to extent the instruments list with more lags

$$z_t = \left(1, \frac{c_t}{c_{t-1}}, \frac{c_{t-1}}{c_{t-2}}, R_t, R_{t-1}\right)',$$

but the results in rows (N7)−(N12) indicate that more lags do not improve the estimates. We could try to improve the model by searching for more instruments, but that is beyond the scope of this example. A second possibility is that the economic model is not a good representation of the data. Some authors have suggested to extend the model to allow habit formation in the Euler equation, but that is also beyond the scope of this section.

A third possibility is that there is not enough variation in the data to identify the shape of the non-linear function in (36). In the data set it holds that $\frac{c_{t+1}}{c_t}$ and R_{t+1} are close to unity. If the variance is small, it holds that

$$\delta \cdot \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \cdot R_{t+1} - 1 \approx \delta \cdot (1)^{-\gamma} \cdot 1 - 1,$$

which is equal to zero with a discount factor of $\delta = 1$ and (virtually) any value for γ.

5 Further Readings

A short and non-technical presentation of the GMM principle and applications in cross-sectional and time series models is given in Wooldridge (2001). The first applications of the methodology are found in Hansen and Singleton (1982) and Hansen and Singleton (1983) both based on a C-CAPM model. Many journal articles use the same framework and according to the Social Sciences Citation Index the first paper is cited more than 500 times.

All presentations of the underlying theory are very technical. The textbook by Hayashi (2000) uses GMM as the organizing principle and the first chapters of that book are readable. The asymptotic theory was first presented in Hansen (1982). The theory of GMM is also covered in the book edited by Mátyás (1999); which also contains many extensions e.g. to non-stationary time series. Technical detail on the estimation of HAC variance matrices are given in Newey and West (1987) and Andrews (1991).

References

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

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

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

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91(2), 249–265.

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

Mátyás, L. (1999): *Generalized Method of Moments Estimation*. Cambridge University Press, Cambridge.

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

VERBEEK, M. (2004): *A Guide to Modern Econometrics*. John Wiley & Sons, 2nd edn.

WOOLDRIDGE, J. M. (2001): "Applications of Generalized Method of Moments Estimation," *Journal of Economic Perspectives*, 15(4), 87–100.