PhD Econometrics - GMM

Professor: Lynda Khalaf - Carleton University

Fall 2012

GMM estimation: motivation

To introduce GMM, let us discuss an application (in fact an extension) of the IV method to more complicated, non-linear models. We start with two examples that illustrate recent dynamic models of rational expectations.

The behavior of agents (e.g. investors, decision makers) is often influenced by their perception (or expectations) of "the future". It is rather difficult to obtain OBSERVED DATA on such expectations.* Nevertheless, one can formulate, and assess, models that aim to capture agents' perception of the future.

A common hypothesis underlying such models is that expec-

* As a counter example, note that some central banks now publish "survey based expectations of inflation".

tations are rational, in the following sense: expectation errors are uncorrelated with all information available at the time agents form their expectations. To set focus, consider the portfolio-choice asset model from Hansen and Singleton (1982).[†]

Let c_t refer to total consumption for an investor at time t; the agent derives utility, denoted, $u(c_t)$, from consuming c_t. At time t, the agent chooses his "consumption strategy", or more formally $\{c_{t+\tau}\}_{\tau=0}^{\infty}$ to maximize his expected total utility function, given the set of information available to him until time t, which is denoted X_t^* :

$$u_t = \sum_{\tau=0}^{\infty} \beta^{\tau} E\left(u\left(c_{t+\tau}\right)|X_t^*\right), \tag{1}$$

where β is a parameter, $0 < \beta < 1$. The smaller β is, the smaller is the weight that the investor attributes to future events, so β may be seen as his "discounting" parameter. Furthermore, at each time t, the agent has the possibility to purchase m different assets, where the returns of asset i at time $t+1$ is $(1 + r_{it+1})$.

† Hansen, Lars P. and Kenneth, J. Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50:1269-86. Errata: *Econometrica* 52:267-68.

The associated budget constraint [for each asset i] is thus given by:

$$A_{it+1} = (1 + r_{it+1})(A_{it} - c_t) \tag{2}$$

where A_{it} is the stock held by the agent from asset i at time t. Given the latter constraint, the consumption strategy of the agent, that is $\{c_{t+\tau}\}_{\tau=0}^{\infty}$, thus solves the problem of maximizing the expected utility function (1) under the constraint (2). The solution to this problem obtains as (the "Euler" condition):

$$u'(c_t) = \beta E\{(1 + r_{it+1}) u'(c_{t+1}) | X_t^*\}, \quad i = 1, 2, ..., m. \tag{3}$$

This condition describes the consumer's choice between current and future consumption. A common form for the utility function $u(c_t)$ is often taken to be:

$$u(c_t) = \begin{cases} \frac{c_t^{1-\gamma}}{1-\gamma}, & \gamma > 0 \text{ and } \gamma \neq 1 \\ \log c_t, & \gamma = 1. \end{cases} .$$

The parameter γ is the coefficient that measures the degree of risk aversion. With this special case, the Euler condition (3) is written as:

$$c_t^{-\gamma} = \beta E \left\{ (1 + r_{it+1})\, c_{t+1}^{-\gamma} | X_t^* \right\}, \quad i = 1, 2, ..., m,$$

and dividing left and right hand side by $c_t^{-\gamma}$:

$$1 = \beta E \left\{ (1 + r_{it+1}) \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} | X_t^* \right\} \tag{4}$$

$$E \left\{ \underbrace{1 - \beta (1 + r_{it+1}) \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma}}_{\epsilon_t(\beta,\gamma)} | X_t^* \right\} = 0. \tag{5}$$

Recall the Law Of Iterated Expectations [Appendix B in textbook]:

$$\underbrace{cov(\varepsilon_t, X_t^*)}_{uncond.\ \mathrm{cov}} = cov_{X_t^*} \left[\epsilon_t (\beta, \gamma), \underbrace{(E (\epsilon_t (\beta, \gamma) | X_t^*))}_{0} \right] = 0$$

It follows that the error $\epsilon_t (\beta, \gamma)$ should be orthogonal (uncorrelated) with any subset of X_t^* :

$$E \left\{ \left[1 - \beta (1 + r_{it+1}) \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \right] X_t \right\} = 0$$

where X_t is the subset of X_t^* that is observed by the analyst; the variables X_t are instruments.

Hansen and Singleton have estimated this model using for c_t real consumption per capita in the US. They considered two assets; and as instruments, they used: a constant term, lags of consumption growth and lags of returns on the two assets considered. To see that the problem is not restricted to financial applications, consider the estimation of a New Keynesian Phillips Curve equation, as proposed in Gali and Gertler (1999) and Gali, Gertler and Lopez-Salido (2001).[‡]

The underlying macro-model is specified as follows. Firms evolve in a monopolistically competitive environment and cannot adjust their prices at all times. A Calvo-type assumption is used to represent the fact that a proportion of firms, θ, do not adjust their prices in period t. In addition, it is assumed that some of the firms do not optimize but use a rule of thumb when setting their prices. The proportion of such firms (referred to as the backward-looking price-setters) is given by ω. In such an envi-

[‡] **Gali J. and M. Gertler (1999)**. *Inflation Dynamics: A Structural Econometric Analysis*, The Journal of Monetary Economics 44, 195-222. **Gali J., Gertler M. and J. D. Lopez-Salido (2001)**. *European Inflation Dynamics*, European Economic Review 45, 1237-1270.

ronment, profit-maximization and rational expectations leads to the following hybrid NKPC equation, where inflation, π_t, is given by

$$\pi_t = \lambda s_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1}, \tag{6}$$

$$\begin{aligned} \lambda &= \frac{(1-\omega)(1-\theta)(1-\beta\theta)}{\theta+\omega-\omega\theta+\omega\beta\theta} \\ \gamma_f &= \frac{\beta\theta}{\theta+\omega-\omega\theta+\omega\beta\theta} \\ \gamma_b &= \frac{\omega}{\theta+\omega-\omega\theta+\omega\beta\theta}. \end{aligned} \tag{7}$$

Here π_{t-1} is the inflation lag, $E_t\pi_{t+1}$ is expected inflation for period $t+1$, and s_t is real marginal costs (expressed as a percentage deviation with respect to its steady-state value). The parameter γ_f is the forward-looking component of inflation, γ_b is its backward-looking part, and β is the subjective discount rate.

Gali and Gertler rewrite the above NKPC model in terms of orthogonality conditions, given by:

$$E_t \left\{ \underbrace{\left(\pi_t - \lambda s_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}\right)}_{\epsilon_t\left(\lambda, \gamma_f, \gamma_b\right)} z_t \right\} = 0$$

where z_t includes variables (instruments) that are orthogonal to $\epsilon_t\left(\lambda, \gamma_f, \gamma_b\right)$.

Quarterly U.S. data are used, with π_t measured by the percentage change in the GDP deflator, and real marginal costs given by the logarithm of the labour income share. Finally, their instruments include four lags of inflation, labour share, commodity-price inflation, wage inflation, the long-short interest rate spread, and output gap (measured by a detrended log GDP). Gali and Gertler's estimations yield values for ω, θ, and β that are (0.27, 0.81, 0.89). Accordingly, Gali and Gertler conclude that the forward-looking component of inflation is more important than the backward-looking part (i.e. the estimated γ_f is larger than the estimated γ_b).

GMM estimation: methodology

To introduce GMM in its simplest form, let us first review linear IV based estimation. For presentation clarity, we will set the following notation. Consider a scalar y_t and a (column) vector x_t, of dimension $N_X \times 1$, and assume the following linear model holds

$$y_t = x_t'\theta_0 + \epsilon_t(\theta_0).$$

Stack the latter as follows

$$\underbrace{X}_{T\times N_X} = \begin{array}{r} \\ \\ x_t' \rightarrow \\ \\ \\ \end{array} \begin{bmatrix} X_{11} & . & X_{1j} & . & X_{1N_X} \\ . & & . & & \\ X_{t1} & . & X_{tj} & . & X_{tN_X} \\ . & & . & & \\ X_{T1} & . & X_{Tj} & . & X_{TN_X} \end{bmatrix}$$

so

$$\underbrace{y}_{T\times 1} = \begin{bmatrix} y_1 \\ . \\ y_t \\ . \\ y_T \end{bmatrix}, \quad \underbrace{x_t}_{N_X\times 1} = \begin{bmatrix} X_{t1} \\ \\ X_{tj} \\ \\ X_{tN_X} \end{bmatrix}, \quad \underbrace{\epsilon(\theta_0)}_{T\times 1} = \begin{bmatrix} \epsilon_1(\theta_0) \\ . \\ \epsilon_t(\theta_0) \\ . \\ \epsilon_T(\theta_0) \end{bmatrix}.$$

This allows to express the model at hand in matrix notation

$$y = \underbrace{X}_{T\times N_X}\theta_0 + \epsilon(\theta_0).$$

Let us first assume that

$$Var\left[\epsilon(\theta_0)\right] = \sigma^2 I_T$$

where I_T is a $T \times T$ identity matrix. You also have a set matrix of N_H instruments,

$$\underbrace{H}_{T\times N_H} = \begin{bmatrix} & H_{11} & . & H_{1j} & . & H_{1N_H} \\ & . & & . & & \\ h_t' \rightarrow & \boxed{H_{t1} \quad . \quad H_{tj} \quad . \quad H_{tN_H}} \\ & . & & . & & \\ & H_{T1} & . & H_{Tj} & . & H_{TN_H} \end{bmatrix}, \quad \underbrace{h_t}_{N_H\times 1} = \begin{bmatrix} H_{t1} \\ . \\ H_{tj} \\ . \\ H_{tN_H} \end{bmatrix}$$

which satisfy the following orthogonality condition:

$$E\left(\underbrace{h_t}_{N_H\times 1}\underbrace{\epsilon_t(\theta_0)}_{1\times 1}\right) = \underbrace{0}_{N_H\times 1}.$$

Given any θ value, let

$$\underbrace{\epsilon_t(\theta)}_{1\times 1} = y_t - x_t'\theta, \quad \underbrace{\epsilon(\theta)}_{T\times 1} = y - X\theta$$

$$\underbrace{f_t(\theta)}_{N_H\times 1} = \underbrace{h_t}_{N_H\times 1}\underbrace{\epsilon_t(\theta)}_{1\times 1}$$

such that

$$E\left(\underbrace{f_t(\theta_0)}_{N_H\times 1}\right) = \underbrace{0}_{N_H\times 1}.$$

IV estimation aims to find an estimate of θ which "matches" this "moment" (i.e. satisfies this condition as much as possible), replacing the expectation by the sample average, thus relying on the law of large numbers. So let

$$\underbrace{g_T(\theta)}_{N_H\times 1} = \frac{1}{T}\sum_{t=1}^{T}\underbrace{f_t(\theta)}_{N_H\times 1} = \frac{1}{T}\sum_{t=1}^{T}\underbrace{h_t}_{N_H\times 1}\underbrace{\epsilon_t(\theta)}_{1\times 1} \tag{8}$$

i.e.

$$\underbrace{g_T(\theta)}_{N_H\times 1}=\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} H_{t1}\epsilon_t(\theta) \\ . \\ H_{tj}\epsilon_t(\theta) \\ . \\ H_{tN_H}\epsilon_t(\theta) \end{bmatrix}=\begin{bmatrix} \frac{1}{T}\sum\limits_{t=1}^{T} H_{t1}\epsilon_t(\theta) \\ . \\ \frac{1}{T}\sum\limits_{t=1}^{T} H_{tj}\epsilon_t(\theta) \\ . \\ \frac{1}{T}\sum\limits_{t=1}^{T} H_{tN_H}\epsilon_t(\theta) \end{bmatrix}$$

or alternatively

$$\underbrace{g_T(\theta)}_{N_H\times 1}=\frac{1}{T}\underbrace{H'}_{N_H\times T}\underbrace{\epsilon(\theta)}_{T\times 1}. \tag{9}$$

Indeed, one can check that

$$\frac{1}{T}\begin{bmatrix} H_{11} & . & \boxed{\begin{matrix} H_{t1} \\ . \\ H_{tj} \\ . \\ H_{tN_H} \end{matrix}} & . & H_{T1} \\ . & & & & \\ H_{1j} & . & & . & H_{Tj} \\ . & & & & \\ H_{1N_H} & . & \underbrace{\phantom{H_{tN_H}}}_{h_t} & . & H_{TN_H} \end{bmatrix}\begin{bmatrix} \epsilon_1(\theta_0) \\ . \\ \epsilon_t(\theta_0) \\ . \\ \epsilon_T(\theta_0) \end{bmatrix}=\begin{bmatrix} \frac{1}{T}\sum\limits_{t=1}^{T} H_{t1}\epsilon_t(\theta_0) \\ \\ \frac{1}{T}\sum\limits_{t=1}^{T} H_{tj}\epsilon_t(\theta_0) \\ \\ \frac{1}{T}\sum\limits_{t=1}^{T} H_{tN_H}\epsilon_t(\theta_0) \end{bmatrix}$$

In general, there are **more** elements in $g_T(\theta)$ than unknown coefficient, so it is not possible to set all the elements of $g_T(\theta)$ to zero. Instead, we minimize the "weighted" sum of squares (and cross-products), i.e. the quadratic form

$$\begin{aligned} Q_T(\theta) &= \underbrace{g_T(\theta)'}_{1\times N_H} \underbrace{W_T}_{N_H\times N_H} \underbrace{g_T(\theta)}_{N_H\times 1} \\ &= \left[\frac{1}{T}\underbrace{\epsilon(\theta)'}_{1\times T}\underbrace{H}_{T\times N_H}\right] \underbrace{W_T}_{N_H\times N_H} \left[\frac{1}{T}\underbrace{H'}_{N_H\times T}\underbrace{\epsilon(\theta)}_{T\times 1}\right] \end{aligned}$$

where W_T is a symmetric, positive definite matrix.

Formally

$$\widehat{\theta}_T = \arg\min_{\theta}\left\{Q_T(\theta)\right\}.$$

In the case where

$$\epsilon(\theta) = y - X\theta$$

so

$$Q_T(\theta) = \left[\frac{1}{T}(y - X\theta)' H\right] W_T \left[\frac{1}{T} H' (y - X\theta)\right]$$

the first order conditions of the latter minimization problem are:

$$\begin{aligned}
\frac{\partial Q_T(\theta)}{\partial \theta} = & \frac{1}{T^2}\frac{\partial (y' H W_T H' y)}{\partial \theta} \\
& -\frac{1}{T^2}\frac{\partial (2\boxed{y' H W_T H' X}\theta)}{\partial \theta} \\
& +\frac{1}{T^2}\frac{\partial (\theta' \boxed{X' H W_T H' X}\theta)}{\partial \theta}
\end{aligned}$$

$$\Rightarrow \boxed{\text{FOC:}} \quad X' H W_T H' y = \boxed{X' H W_T H' X}\,\widehat{\theta}_T$$

$$\boxed{\widehat{\theta}_T = \left[X' H W_T H' X\right]^{-1} X' H W_T H' y}$$

When N_H equals the number of parameters, $H'X$ becomes a square invertible matrix so

$$\begin{aligned}
\widehat{\theta}_T &= (X'HW_TH'X)^{-1}X'HW_TH'y \\
&= (H'X)^{-1}(W_T)^{-1}(X'H)^{-1}X'HW_TH'y \\
&= (H'X)^{-1}(W_T)^{-1}\underbrace{(X'H)^{-1}X'H}W_TH'y \\
&= (H'X)^{-1}\underbrace{(W_T)^{-1}W_T}H'y \\
&= (H'X)^{-1}H'y.
\end{aligned}$$

Regular Asymptotic Theory:

$$
\begin{aligned}
\widehat{\theta}_T &= \left[X'HW_TH'X\right]^{-1} X'HW_TH'(X\theta_0 + \epsilon(\theta_0)) \\
&= \left[X'HW_TH'X)\right]^{-1} X'HW_TH'X \quad \theta_0 \\
&\quad + \left[X'HW_TH'X)\right]^{-1} X'HW_TH'\epsilon(\theta_0)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\theta}_T - \theta_0 &= \left[X'HW_TH'X\right]^{-1} X'HW_TH'\epsilon(\theta_0) \\
&= \left[\frac{X'H}{T}W_T\frac{H'X}{T}\right]^{-1} \frac{X'H}{T}W_T\frac{H'\epsilon(\theta_0)}{T}
\end{aligned}
$$

so that

$$
\sqrt{T}(\widehat{\theta}_T - \theta) = \left[\frac{X'H}{T}W_T\frac{H'X}{T}\right]^{-1} \frac{X'H}{T}W_T\frac{H'\epsilon(\theta_0)}{\sqrt{T}}.
$$

Observe that

$$
\frac{H'\epsilon(\theta_0)}{\sqrt{T}} = \sqrt{T}g_T(\theta_0) = \boxed{\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t(\theta_0)} \blacktriangleleft\blacktriangleleft
$$

so for $T \rightarrow \infty$

$$
\begin{aligned}
\frac{X'H}{T} &\rightarrow M_{XH} \text{ and } \frac{H'X}{T} \rightarrow M_{HX=}M'_{XH} \\
W_T &\rightarrow W \\
\frac{H'H}{T} &\rightarrow M_{HH}
\end{aligned}
$$

$$\frac{H'\epsilon(\theta_0)}{\sqrt{T}} \overset{asy}{\sim} N(0,S)$$

$$\begin{aligned} S &= \lim_{T\to\infty} Var\left[\frac{H'\epsilon(\theta_0)}{\sqrt{T}}\right] \\ &= \lim_{T\to\infty} E\left[\frac{H'\epsilon(\theta_0)\,\epsilon(\theta_0)'\,H}{T}\right] = \sigma^2 M_{HH}. \end{aligned}$$

For further reference, given (8)-(9) it is worth noting that

$$\boxed{S = \lim_{T\to\infty} Var\left[\sqrt{T} g_T(\theta)\right] = \lim_{T\to\infty} Var\left[\tfrac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t(\theta)\right]} \blacktriangleleft\blacktriangleleft$$

$$= \lim_{T\to\infty} Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \underbrace{h_t}_{N_H\times 1}\underbrace{\epsilon_t(\theta)}_{1\times 1}\right].$$

It follows that

$$\sqrt{T}(\widehat{\theta}_T - \theta) \overset{asy}{\sim} N(0,V)$$

$$V =$$

$$[M_{XH}\ (W)\,M_{HX}]^{-1}\,M_{XH}\,(W)\ (S)\,(W)\,M_{HX}\ [M_{XH}\ (W)\,M_{HX}]^{-1}.$$

It is possible to consider a weighting matrix W_T which minimizes the asymptotic variance of the IV estimator: it is obvious the latter minimum is achieved when

$$\begin{aligned} W &= (\textit{any} \text{ positive scalar})\ S^{-1} \\ &= (\textit{any} \text{ positive scalar})\ \left(\sigma^2 M_{HH}\right)^{-1} \end{aligned}$$

so for convenience, one may pick the optimal weight

$$W^* = M_{HH}^{-1}$$

so the asymptotic variance of the IV estimator reduces to

$$V^* = \sigma^2 \left[M_{XH}\ M_{HH}^{-1} M_{HX}\right]^{-1}.$$

So the matrix W_T^* which should converge to W^* may be considered as follows

$$W_T^* = \left(\frac{H'H}{T}\right)^{-1}$$

which leads to the IV estimator

$$\boxed{\widehat{\theta}_T = \left[X'H\left(H'H\right)^{-1}H'X\right]^{-1} X'H\left(H'H\right)^{-1}H'y}$$

and an estimate of its covariance matrix follows naturally:

$$V_T^* = \widehat{\sigma}^2(X'H(H'H)^{-1}H'X)^{-1}$$

where $\widehat{\sigma}^2$ is a consistent estimate of the variance of $\epsilon_t(\theta_0)$, for example

$$\widehat{\sigma}^2 = \frac{1}{T}\epsilon(\widehat{\theta}_T)'\epsilon(\widehat{\theta}_T).$$

The latter estimator can be expressed in the 2SLS form, namely:

- ▶ regress X on H and save $\hat{X} = H(H'H)^{-1}H'X$
  - note that

$$\begin{aligned}\hat{X}'\hat{X} &= X'H(H'H)^{-1}H'H(H'H)^{-1}H'X \\ &= X'H(H'H)^{-1}H'X\end{aligned}$$

  ▶ regress y on $\hat{X}$ and obtain

$$\begin{aligned}\widehat{\theta}_T &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\ &= (X'H(H'H)^{-1}H'X)^{-1}X'H(H'H)^{-1}H'y\end{aligned}$$

- It is worth noting that N_H equals the number of parameters so that $H'X$ becomes a square invertible matrix so that

$$\widehat{\theta}_T = (H'X)^{-1} H'y$$

and if we replace this expression in the objective function which corresponds to W_T^*, then

$$\widehat{Q}_T\left(\widehat{\theta}_T\right) =$$

$$\frac{1}{T^2}\left(y - X\boxed{\left(H'X\right)^{-1}H'y}\right)' H(H'H)^{-1}\underbrace{H'\left(y - X\boxed{\left(H'X\right)^{-1}H'y}\right)}$$

$$= 0.$$

The minimum in this case thus occurs at zero..

Let us now extend these arguments in several directions. Suppose we have a model which defines a VECTOR of dimension $N_\epsilon \times 1$ rather than a scalar

$$\underbrace{\epsilon_t}_{N_\epsilon \times 1} = \epsilon(x_t, \theta)$$

where x_t now includes all data available (*i.e.* we drop the distinction between y and x), the dimension of θ is $N_\beta \times 1$, the function $\epsilon(x_t, \theta)$ may be non-linear in variables and in parameters, and ϵ_t can be heteroskedastic and serially correlated. We retain the previous definition of the matrix of instrument, and the model only implies the following.

Setting

$$\underbrace{f_t(\theta)}_{N_\epsilon N_H \times 1} = \underbrace{h_t}_{N_H \times 1} \otimes \underbrace{\epsilon(x_t, \theta)}_{N_\epsilon \times 1}$$

i.e. every element in h_t will be multiplied by $\epsilon(x_t, \theta)$, the model implies that

$$E(f_t(\theta_0)) = 0.$$

We then proceed as in the linear case (it is important at this stage to track the dimension of the matrices in question):

$$\underbrace{g_T(\theta)}_{N_H N_\epsilon \times 1} = \frac{1}{T}\sum_{t=1}^{T} \underbrace{f_t(\theta)}_{N_H N_\epsilon \times 1} = \frac{1}{T}\sum_{t=1}^{T} \underbrace{h_t}_{N_H \times 1} \otimes \underbrace{\epsilon_t(\theta)}_{N_\epsilon \times 1}$$

which leads to the estimator

$$\begin{aligned}
\widehat{\theta}_T &= \arg\min_{\theta} \{Q_T(\theta)\} \\
Q_T(\theta) &= \underbrace{g_T(\theta)'}_{1\times(N_H N_\epsilon)} \underbrace{W_T}_{(N_H N_\epsilon)\times(N_H N_\epsilon)} \underbrace{g_T(\theta)}_{(N_H N_\epsilon)\times 1} .
\end{aligned}$$

Of the course, the minimization has to be performed numerically.

The FOC are:

$$D_T\left(\widehat{\theta}_T\right)' \underbrace{W_T\left[g_T\left(\widehat{\theta}_T\right)\right]}_{N_H N_\epsilon \times 1} = 0$$

$$\underbrace{D_T\left(\widehat{\theta}_T\right)}_{N_H N_\epsilon \times N_\beta} = \frac{\partial g_T(\theta)}{\partial \theta'}.$$

So defined,

$$D_T\left(\widehat{\theta}_T\right) \to D_0 = E\left(\frac{\partial f_t(\theta)}{\partial \theta'}\right),$$

and let

$$\boxed{S = \lim_{T\to\infty} Var\left[\sqrt{T} g_T(\theta)\right] = \lim_{T\to\infty} Var\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t(\theta)\right]} \blacktriangleleft\blacktriangleleft$$

as in the linear case. Then it can be shown that:

$$\sqrt{T}(\widehat{\theta}_T - \theta) \overset{asy}{\sim} N(0, V)$$

$$V =$$

$$\left[D_0' \; (W) D_0\right]^{-1} D_0' (W) \; (S) (W) D_0 \; \left[D_0' \; (W) D_0\right]^{-1}$$

which suggests the following choice for an optimal weighting matrix

$$W^* = S^{-1}$$

and leads to the optimal asymptotic variance

$$V^* = \left[D_0'\, S^{-1} D_0\right]^{-1}.$$

The following practical steps can be followed to estimate these matrices.

1. Chose an identity matrix (for example) as a weighting matrix and minimize the objective function to obtain an initial estimator say $\widehat{\theta}_T^0$.
2. Using the latter, estimate D_0 using $D_T\left(\widehat{\theta}_T^0\right)$ and obtain an estimate of S [see below], depending on whether errors are assumed to be iid or not. These estimates will provide a "round two" weighting matrix.
3. Plug this weighting matrix in the objective function and iterate.

$$\boxed{\text{On the estimation of } S}$$

$$\begin{aligned} S &= \lim_{T\to\infty} Var\left[\sqrt{T} g_T(\theta_0)\right] = \lim_{T\to\infty} Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t(\theta_0)\right] \\ &= \lim_{T\to\infty} E\left[\frac{1}{T}\left(\sum_{t=1}^{T} f_t(\theta_0)\right)\left(\sum_{t=1}^{T} f_t(\theta_0)'\right)\right]. \end{aligned}$$

In the $i.i.d.$ case, the following estimate may be used

$$\hat{S}(\theta) = \frac{1}{T}\sum_{t=1}^{T} f_t(\theta) f_t(\theta)'$$

where we emphasize its dependence on θ. Since GMM is often used in the non-iid case, an estimator such as the Newey-West one can be used:

$$\begin{aligned} \hat{S}(\theta) &= \hat{\Gamma}_0(\theta) + \sum_{j=1}^{q}\left(\frac{q-j}{q}\right)\left(\hat{\Gamma}_j(\theta_0) + \hat{\Gamma}_j'(\theta_0)\right) \\ \hat{\Gamma}_0(\theta) &= \frac{1}{T}\sum_{t=j+1}^{T} f_t(\theta_0) f_{t-j}(\theta_0)'. \end{aligned}$$

This is the same estimator applied for the Newey-West estimator of $\widehat{V}(\widehat{\beta}_{OLS})$ in the linear regression model:

$$
\begin{aligned}
\widehat{V}(\widehat{\beta}_{OLS}) &= (X'X)^{-1}\boxed{\widehat{Q}}(X'X)^{-1} \\
\widehat{Q} &= \sum_{i=1}^{n}\widehat{\varepsilon}_i^2 x_i x_i' + \sum_{l=1}^{L}\sum_{i=l+1}^{n} w_l \widehat{\varepsilon}_i \widehat{\varepsilon}_{i-l}\left(x_i x_{i-l}' + x_{i-l} x_i'\right) \\
w_l &= 1 - \frac{l}{L+1}
\end{aligned}
$$

Exercises

Question 1

A prototypical New Keynesian Phillips Curve [NKPC] takes the form

$$\pi_t = \lambda s_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \eta_t, \quad t = 1, ..., T \tag{10}$$

where π_t is inflation, s_t is a driving variable [assumed endogenous] measured via the output gap, unemployment or real marginal costs and η_t is an unobserved shock. For estimation purposes, rational expectations are imposed

$$\pi_{t+1} = E_t \pi_{t+1} + \upsilon_{t+1} \tag{11}$$

leading to the estimating equation

$$\pi_t = \lambda s_t + \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + \epsilon_t, \quad t = 1, ..., T \tag{12}$$

where ϵ_t includes expectation errors. Macroeconomic theory also supports versions of the curve that exclude the η_t shock, of the form

$$\pi_t = \lambda s_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1}, \quad t = 1, ..., T \tag{13}$$

in which case rational expectations introduce expectation errors into the estimating equation via (11).

1. Explain how the GMM method can be applied to estimating each of (10) and (13) imposing rational expectations.[§] Explain how instruments can be introduced into your proposed objective function.

2. Suppose that in the context of (10), we aim to estimate ϑ and ϕ rather than γ_f and γ_b, with the same data and instrument sets where

$$\gamma_f = \frac{\phi}{1 + \phi\vartheta}, \quad \gamma_b = \frac{\vartheta}{1 + \phi\vartheta}. \tag{14}$$

The parameter ϑ captures the extent of indexation to past inflation and ϕ is the subjective discount rate. Explain how this estimation can be done using GMM and emphasize in what way your proposed method differs from your answer to sub-question 1. Specifically, suppose the GMM estimators you obtain are de-

§ Suggestion: study the time series properties of the error term in (12).

noted $\hat{\vartheta}$ and $\hat{\phi}$. Now define

$$\hat{\gamma}_f = \frac{\hat{\phi}}{1 + \hat{\phi}\hat{\vartheta}}, \quad \hat{\gamma}_b = \frac{\hat{\vartheta}}{1 + \hat{\phi}\hat{\vartheta}}. \tag{15}$$

Do you expect $\hat{\gamma}_f$ and $\hat{\gamma}_b$ to numerically coincide with the output of the GMM estimator you introduced as an answer to sub-question 1 if the same data and instruments sets are used? Justify your answer.

Question 2.

Consider the linear regression model

$$Y = X\beta + U \tag{16}$$

where $Y = [y_1, \ ... \ , \ y_T]'$, X is $T \times k$, of full column rank, β is $k \times 1$ (unknown) and $U = [u_1, \ \ldots \ , \ u_T]'$ is the vector of error terms of variance σ^2. You also have a set of h instruments; let Z (of dimension $T \times h$) the matrix of instruments.

1. Express the OLS estimator of β in the form of a GMM estimator.

2. Consider the estimator of β obtained by minimizing the quadratic form

$$(Y - X\beta)'Z\,(Z'Z)^{-1}\,Z'(Y - X\beta). \tag{17}$$

Show (justify your answer) that:

(a) this estimator is of the GMM form;

(b) for the exact-identified case, the minimum of expression (17) is equal to zero and is attained with the usual IV estimator

(c) for the over-identified case, show that the weighting matrix $Z'Z$ (in the quadratic form (17)) is proportional to the variance/covariance matrix of the vector of underlying moments.

Question 3.

A GMM estimator $\hat{a}_n$ (of a parameter a) based on the data set Y_i, X_i, $i = 1, .., n$, and a weight matrix S_n, is typically defined as follows:

$$\hat{a}_n = \arg\min_a \left[\sum_{i=1}^{n} h(Y_i, X_i; a)\right]' S_n \left[\sum_{i=1}^{n} h(Y_i, X_i; a)\right]$$

given relevant conditions for the function $h(.; a)$ and the matrix S_n.

1. Show how a moment-matching argument is reflected in this objective function
2. How can instruments intervene in the above definition?
3. Is it possible that the minimum of the objective function is equal to zero? If yes, give an example, with proofs.
4. Is it possible that the minimum of the objective function corresponds to an OLS estimator? If yes, give an example, with proofs.

Question 4.

1. Let $\epsilon(x_t, \theta)$ refer a vector of Euler errors of dimension $N_\epsilon \times 1$ where x_t includes all data available, θ is an unknown parameter of dimension of is $N_\theta \times 1$, and $\epsilon(x_t, \theta)$ may be non-linear

in variables and in parameters, and can be heteroskedastic and serially correlated. Setting $f_t(\theta) = h_t \otimes \epsilon(x_t, \theta)$, consider a model defined by

$$E(f_t(\theta_0)) = 0.$$

2. Define the GMM estimator associated with this model and provide an expression for its asymptotic variance

3. Suppose that the model is linear in variables and in parameters; explain how this will affect the formulas you provided for the previous two sub-question. In particular, would you have analytical solutions for the point estimator and the standard errors? If you need further assumptions to obtain analytical solutions, provide them explicitly.

4. Show how this estimator can be applied to the inflation model

$$\pi_t = \lambda s_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1}, \tag{18}$$

$$\begin{aligned}
\lambda &= \frac{(1-\omega)(1-\theta)(1-\beta\theta)}{\theta+\omega-\omega\theta+\omega\beta\theta} \qquad (19)\\
\gamma_f &= \frac{\beta\theta}{\theta+\omega-\omega\theta+\omega\beta\theta}\\
\gamma_b &= \frac{\omega}{\theta+\omega-\omega\theta+\omega\beta\theta}.
\end{aligned}$$

In particular, explain how you would proceed if you intend to estimate the "reduced form" parameters λ, γ_f and γ_b only rather than the "deep parameters" ω, β and θ, and if you could assume i.i.d. errors.

Question 5.

GMM estimation is motivated by the Law of Large numbers. Suppose that

$$Y_t = \mu + \rho Y_{t-1} + u_t, \qquad t = 1, ..., T,$$

where

$$\begin{aligned}
E(u_t) &= 0, \qquad V(u_t) = \sigma^2, \qquad \forall t \qquad (20)\\
cov(u_j, u_s) &= 0 \qquad \forall j \neq s. \qquad (21)
\end{aligned}$$

Evaluate the following propositions [True, False, Uncertain; **proof required**]:

1. Since the series Y_t , $t = 1, ..., T$ is not *i.i.d.*, then the law of large numbers does not apply. If you need further assumptions to answer this question, provide them explicitly.

$$\left[\text{PLIM} \frac{1}{T} \sum_{t=1}^{T} Y_t \right] \neq E(Y_t).$$

2. Now assume that

$$\begin{aligned} u_t &= \varepsilon_t h_t, \quad t = 1, ..., T, \quad h_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2, \qquad (22) \\ \varepsilon_t &\sim N(0,1) \text{ and } \varepsilon_t \text{ is independent from } h_t, \qquad (23) \end{aligned}$$

   and assess the same proposition for $\rho = 0$ and $\rho \neq 0$. If you need further assumptions to answer this question, provide them explicitly.