# GMM estimation

Fabio Canova
EUI and CEPR
September 2014

## Outline

- GMM and GIV.
- Choice of weighting matrix.
- Constructing standard errors of the estimates.
- Testing restrictions.
- Examples.

## References

Hamilton, J., (1994), *Applied Time Series Analysis*, Princeton University Press.

Hayashi, F. (2002), *Econometrics*, Princeton University Press.

Gali, J and Gertler, M. (1999), "Inflation Dynamics: A Structural Econometric Analysis, *Journal of Monetary Economics*, 44, 195-222.

Hansen, L.P., (1982), "Large Sample Properties of GMM Estimators", *Econometrica*, 50, 1029-1054.

Hansen, L.P. and Singleton, K., (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-1286. (corrigenda, 1984).

Hansen, L. and Hodrick, R. (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis, *Journal of Political Economy*, 88, 829-853.

Hall, A., (1992), "Some Aspects of Generalized Method of Moment Estimators", in G.S. Maddala, C.R. Rao and H.D. Vinod (eds.), *Handbook of Statistics*, vol. 11, Elsevier Science.

Hall, A. and Sen, A. (1999) "Structural Stability testing in models estimated by GMM", *Journal of Business and Economic Statistics*, 17(3), 335-348.

Harding, D. and Pagan, A. (2006) "Synchronization of cycles", *Journal of Econometrics*, 132(1), 59-79.

Ogaki, M., (1993), "GMM: Econometric Applications", in G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., *Handbook of Statistics*, vol. 11, Elsevier Science.

Newey, W. (1990) "Efficient Instrumental variable estimation of nonlinear models", *Econometrica*, 58, 809-837.

Newey, W. and West, K., (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.

Pagan, A. and Yoon, Y., (1993), "Understanding Some Failures of Instrumental Variable Estimators", University of Rochester, manuscript.

Christiano, L. and den Haan, W. (1996) "Small Sample Properties of GMM for Business Cycle Analysis", *Journal of Business and Economic Statistics*, 14, 309-327.

Burnside, C. and Eichenbaum, M. (1996) "Small Sample properties of GMM-based Wald Tests", *Journal of Business and Economic Statistics*, 14, 294-308.

Furher, J. Moore, J. and Schuh, S. (1995) "Estimating the Linear Quadratic Inventory model, ML vs GMM", *Journal of Monetary Economics*, 35, 115-157.

# 1 Introduction

• Dynamic rational expectation models are difficult to estimate:

- There are expectations of future variables.

- The decision rules are often non-linear in the parameters.

- If they are of general equilibrium type, need to solve the model (to obtain a "final form") and use (full information) maximum likelihood techniques.

- Can we estimate the parameters without computing a final form?

- Yes, there is a class of *distance* estimators which can be used to estimate the parameters of a structural model using just the first order conditions.

- These estimators can be used also in many other situations.

- Very general econometric framework

## 2 Definition of GMM estimator

- GMM minimizes distance between sample and population functions.

If there are $m$ conditions, we need weights to transform the vector of $m$ conditions into an index.

- $g_{\infty}(\theta) \equiv$ vector of population (theoretical) functions.

- $g_T(\theta) \equiv$ vector of sample functions (sample size is T).

$$\theta_{GMM} = \text{argmin}[g_T(\theta)' W_T g_T(\theta) - g_{\infty}(\theta)' W g_{\infty}(\theta)] \qquad (1)$$

where $W_T$ is a $m \times m$ full rank matrix, converging in probability to some $W$.

In many economic models the $g$ function are "orthogonality" conditions:

$$g_\infty(\theta) = E_t[g(y_t, \theta) - C] = 0$$

where $\theta$ is a set of parameters, $y_t$ are observable data, $C$ a constant and $E_t(.)$ denotes conditional expectation operators. Hence GMM solves:

$$\theta_{GMM} = \text{argmin}[g_T(\theta)' W_T g_T(\theta)] \tag{2}$$

## 2.1 Examples of economic orthogonality conditions

**Example 2.1** *Model of consumption and savings.*

$$\max_{\{c_t, sa_{t+1}\}_{t=0}^{\infty}} E_0 \sum_t \beta^t u(c_t) \tag{3}$$

$$c_t + sa_{t+1} \leq w_t + (1+r)sa_t \tag{4}$$

$w_t$ *= labor income,* $sa_{t+1}$ *= savings, and* $r_t = r$ $\forall t$*. Letting* $U_{c,t} = \partial u(c_t)/\partial c_t$*, and* $\lambda_t$ *the Lagrangian on the constraint at* $t$*, the FOC are*

$$U_{c,t} = \lambda_t \tag{5}$$

$$\lambda_t = \beta E_t \lambda_{t+1}(1+r) \tag{6}$$

*Combining the two equations leads to the Euler equation:*

$$E_t[\beta(1+r)\frac{U_{c,t+1}}{U_{c,t}} - 1] = 0 \tag{7}$$

*(7) is an orthogonality condition for* $g(y_t, \theta) = [\beta(1+r)U_{c,t+1}/U_{c,t}]$ *and* $C = 1$.

Suppose $u(c_t) = \ln c_t$. If we assume that $r$ is given, (7) is one condition in one unknown parameter $\beta$ (just-identified case).

**Example 2.2** *Lucas Asset Pricing Model (endogenous $r_t$).*

*Agents choose $\{c_t, B_{t+1}, S_{t+1}\}_{t=0}^{\infty}$ to maximize $E_0 \sum_t \beta^t u(c_t)$ subject to*

$$c_t + B_{t+1} + p_t^s S_{t+1} \leq w_t + (1 + r_t) B_t + (p_t^s + sd_t) S_t \tag{8}$$

*$B_t(S_t)$ are bond (stock) holdings , $sd_t$ dividends and $p_t^s$ stock prices. Optimality implies:*

$$\begin{aligned} E_t[\beta \frac{U_{c,t+1}}{U_{c,t}} \frac{(p_{t+1}^s + sd_{t+1})}{p_t^s} - 1] &= 0 \\ E_t[\beta \frac{U_{c,t+1}}{U_{c,t}} - \frac{1}{1 + r_t}] &= 0 \end{aligned} \tag{9}$$

*(9) are two orthogonality conditions for* $g_1(y_t, \theta) = \beta \frac{U_{c,t+1}}{U_{c,t}} \frac{p^s_{t+1} + sd_{t+1}}{p^s_t}$ *, and* $g_2(y_t, \theta) = \beta(1 + r_t)\frac{U_{c,t+1}}{U_{c,t}}$ *and* $C = 1$*.*

If $u(c_t) = \ln c_t$, (9) are two conditions in one unknown parameter $\beta$. (over-identified case). How do we find $\beta$ in this case?

- Could use either one of the two conditions.

- Could also combine assigning weights $w_1$ and $w_2$, i.e. $\min_\beta (w_{1T} g^2_{1T}(\beta) + w_{2T} g^2_{2T}(\beta))$. In this case estimation becomes more efficient (we use all available information).

**Example 2.3** *Real Business Cycle Model*

$$\max_{\{c_t, N_t, K_{t+1}\}_{t=0}^{\infty}} E_0 \sum_t \beta^t u(c_t, N_t) \tag{10}$$

$$c_t + K_{t+1} \leq f(K_t, N_t, \zeta_t) + (1-\delta)K_t \tag{11}$$

$N_t =$ *hours,* $0 \leq N_t \leq 1, K_t =$ *capital and* $\zeta_t =$ *technology disturbance. The Euler equation for capital accumulation is*

$$E_t(\beta \frac{U_{c,t+1}}{U_{c,t}}[f_{K,t+1} + (1-\delta)] - 1]) = 0 \tag{12}$$

*where* $U_{c,t} = \frac{\partial u(c_t, N_t)}{\partial c_t}$, $f_{K,t+1} = \frac{\partial f}{\partial K_t}$. *(12) is an orthogonality condition for* $g(y_t, \theta) = \beta \frac{U_{c,t+1}}{U_{c,t}}(f_K + (1-\delta))$ *and* $C = 1$.

If $u(c_t) = \ln c_t + V(1 - N_t)$, we have one condition, and at least two parameters $(\beta, \delta)$ to estimate (under-identified case). Unless we add other orthogonality conditions, we can not estimate both parameters.

**Conclusion: Dynamic models where agents have rational expectations produce orthogonality conditions as the result of optimization. Parameters entering these conditions can be estimated with GMM.**

How do we add conditions if we are as in a situation like example 2.3?

- Use the law of iterated expectations: $E(E_t g(y_t, \theta)) = Eg(y_t, \theta) = 0$. Hence, we can substitute conditional expectations $E_t$ with unconditional ones. This extends the set of conditions to which GMM can be applied. In fact:

• GMM be applied to static optimality conditions, e.g. in example 2.3 one extra condition is $\frac{U_{N,t}}{U_{c,t}} = f_N$. Since this holds for every $t$, it must be that it holds on average, i.e. $E(\frac{U_{N,t}}{U_{c,t}} - f_N) = 0$.

• GMM can be applied to identities: e.g. budget constraint $c_t + k_{t+1} - (1-\delta)k_t = y_t$ implies $E(c_t + k_{t+1} - (1-\delta)k_t - y_t) = 0$.

## 2.2 Definition of GIV estimators

• In example 2.3 there are more parameters than orthogonality conditions. We could jointly estimate the Euler equation and the intratemporal conditions if $\beta, \delta$ entered there but they don't. How do we proceed in this case?

Let $z_t$ be any variables which is observable at time $t$. If $E_t[g(y_t, \theta) - C] = 0$ then $E\{[g(y_t, \theta) - C]z_t\} = 0$ and

$$\theta_{GIV} = \text{argmin}[z_T' g_T(\theta)' W_T g_T(\theta) z_T] \qquad (13)$$

In example 2.3, let $u(c_t) = \ln c_t$, then:

$$E_t(\beta \frac{c_t}{c_{t+1}}[f_{K,t+1} + (1-\delta)] - 1]) = 0 \tag{14}$$

If $c_{t-1}, f_{K,t}$ belong to the information set of agents, to estimate $\beta, \delta$ we could use

$$\begin{aligned} E(\beta \frac{c_t}{c_{t+1}}[f_{K,t+1} + (1-\delta)] - 1])\frac{c_{t-1}}{c_t} &= 0 \qquad (15) \\ E(\beta \frac{c_t}{c_{t+1}}[f_{K,t+1} + (1-\delta)] - 1])f_{K,t} &= 0 \qquad (16) \end{aligned}$$

- If $u(c_t)$ has additional parameters, we could use $\frac{c_{t-2}}{c_{t-1}}, f_{k,t-1}$, as instruments to "increase" the number of orthogonality conditions.

- How do we choose $z_t$? Difficult!

- How many $z_t$ should I use? If, for example, I add to the above

$$E(\beta \frac{c_t}{c_{t+1}}[f_{K,t+1} + (1-\delta)] - 1])\frac{c_{t-2}}{c_{t-1}} = 0$$

We now have three equations in two unknowns. Need to use a weighting matrix whenever dim(z)$> dim(\theta)$.

## 2.3 Examples of econometric orthogonality conditions

- OLS: $E(x_t'e_t) = 0$. Then $g(x_t, \theta) = (x_t'y_t - x_t'x_t'\beta) = x_t'e_t$ and $C = 0$.

- IV: $E(z_t'e_t) = 0$. Then $g(x_t, z_t, \theta) = (z_t'y_t - z_t'x_t'\beta) = z_t'e_t$ and $C = 0$.

- ML: $\mathcal{L} = \prod_t f(y_t, \theta)$; $\frac{\partial \ln \mathcal{L}(\theta_{ML})}{\partial \theta} = 0$, $g(x_t, \theta) = \frac{\partial \ln f_t(y_t, \theta)}{\partial \theta}$ and $C = 0$.

**Conclusion: linear models featuring (unconditional) orthogonality conditions among the identifying assumptions can be estimated with GMM.**

# 3 Mechanics of GMM in linear models

- Model: $y_t = \beta x_t + e_t, \quad E(e_t|x_t) \neq 0, \; E(e_t e_t') = \sigma^2$.

- We have available $z_t$ such that $E_t(e_t|z_t) = 0, E_t(x_t|z_t) \neq 0$. Want:

$$\min_{\beta} Q_T(\beta) = (e(\beta)'z)W_T(e(\beta)'z)'$$

where $e(\beta)'z \equiv \frac{1}{T}\sum_t(e_t(\beta)'z_t)$ and it is assumed to converge to $E(e_t z_t) = 0$ as $T$ becomes large, and $W_T$ is a weighting matrix. The F.O.C. are:

$$x_T' z_T W_T z_T' y_T = x_T' z_T W_T z_T' x_T \beta \tag{17}$$

Two cases:

a) If dim($\beta$) $=$ $dim(z)$; $z_T' x_T W_T$ is a square matrix and $\beta_{GMM} = (z_T' x_T)^{-1} z_T' y_T$. This is the standard IV estimator.

b) If dim($\beta$) $< dim(z)$; $\beta_{GMM} = (x_T' z_T W_T z_T' x_T)^{-1} (x_T' z_T W_T z_T' y_T)$.

This is because $z_T' x_T W_T$ is rectangular and $z_T' x_T W_T Z_T e_T = 0$ does not imply $Z_T e_T = 0$ (only $dim(\beta)$ conditions are set to zero, $dim(z) - dim(\beta)$ are unrestricted).

**- $z$ must correlated with $x$ otherwise $(z_T' x_T)^{-1}$ may not be computable and variance of estimator may be large.**

Two results:

• GMM estimators are consistent under general conditions, i.e. as $T \rightarrow \infty$, $\beta_{GMM} \rightarrow \beta$ with probability one.

• GMM estimators are asymptotically normal i.e. $T^{0.5}(\beta_{GMM} - \beta) \xrightarrow{D} \mathbb{N}(0, \Sigma_\beta)$ where

$$\begin{aligned} \Sigma_\beta &= \Sigma_{z,x}^{-1} \sigma^2 \Sigma_{z,z} \Sigma_{x,z}^{-1} \quad \text{case a)} && (18) \\ &= (\Sigma_{x,z} W \Sigma_{z,x})^{-1} (\Sigma_{x,z} W \sigma^2 \Sigma_{z,z} W \Sigma_{z,x}) (\Sigma_{x,z} W \Sigma_{z,x})^{-1'} \quad \text{case b)} && (19) \end{aligned}$$

• Covariance matrix in (19) depends on $W$. How to choose W? Choose $W$ to minimize $\mathbf{\Sigma}_\beta$ (asymptotic efficiency).

Solution: $\hat{W} = \sigma^{-2}\mathbf{\Sigma}_{z,z}^{-1}$. Then

$$\mathbf{\Sigma}_\beta^* \equiv \mathbf{\Sigma}(\hat{W}) = \mathbf{\Sigma}_{z,x}^{-1}\sigma^2\mathbf{\Sigma}_{z,z}\mathbf{\Sigma}_{x,z}^{-1'} \quad (20)$$

$$\beta_{GMM}^* \equiv \beta(\hat{W})_{GMM} = (x_T'\hat{x}_T)^{-1}(\hat{x}_T'y_T) \quad (21)$$

where $\hat{x}_T = z_T'(z_T'z_T)^{-1}z_T'x_T$.

Implications:

i) Optimal W is proportional to the covariance matrix of $z$ $\longrightarrow$ use as weights the relative variability of instruments.

ii) $\beta_{GMM}$ with the optimal $W$ is nothing else than 2SLS.

If the model is nonlinear in variables and parameters what can we say about the properties of GMM estimators?

- In nonlinear model GMM is also consistent and asymptotically normal

To obtain this result we need:

i) $y_t$ must be covariance stationary and ergodic.

ii) The function $g$ must be martingale difference, e.g. $E_t[g_t|\mathcal{F}_t] = 0$.

iii) $E(g_t(y_t, \theta) - C) = 0$ must have a unique solution.

iv) A set of technical conditions (e.g. true parameter is not on the boundary of parameter space; space of $\theta$ is compact, etc.).

## Summary of the results

a) Just identified case ($dim(g) = dim(\theta)$):

- $\theta_{GMM} \xrightarrow{P} \theta_0$.

- $\theta_{GMM} \xrightarrow{D} N(\theta_0, T^{-1}B^{-1}AB^{-1'})$

where $B = E_t\frac{\partial g(\theta_0)}{\partial \theta_0}, A = E_t(g_t(\theta_0)g_t(\theta_0)')$.

Since A,B are unknown, we can use consistent estimators in the formulas.

Test of hypothesis are standard, i.e. for a t-test on $\theta_{1,GMM}$ use $(\theta_{1,GMM} - \bar{\theta}_1)/(T^{-1}B^{-1}AB^{-1'})^{0.5}_{11}$ and compare it to a t-distribution table.

b) Overidentified case ($dim(g) > dim(\theta)$). Assume $W_T$ converges to some fixed $W$ a $m \times m$ full rank matrix as $T \to \infty$. Then:

- $\theta_{GMM} \xrightarrow{P} \theta_0$

- $\theta_{GMM} \xrightarrow{D} N(\theta_0, T^{-1}B^{-1}AB^{-1\prime})$

if $W_T$ is any matrix converging to $A^{-1}$ as $T \to \infty$ (this is the optimal weighting matrix in the non-linear case).

## Implementation algorithm

In the overidentified case, we need to know $W$ to construct $\theta_{GMM}$ and we need $\theta_{GMM}$ to construct $W$. So we need an iterative approach. Assume $g_\infty = 0$

1) Set $W_T^0 = I$.

2) Solve $\theta_{GMM}^1 = \text{argmin}[g_T(\theta)' W_T^0 g_T(\theta)]$

3) Set $W_T^1 = T \times [\sum_t g_t(\theta_{GMM}^1) g_t(\theta_{GMM}^1)']^{-1}$.

4) Repeat steps 2 and 3. until $||W_T^i - W_T^{i-1}|| < \iota$, $\iota$ small,$||\theta^i - \theta^{i-1}|| < \iota$, or both.

When iterations are over, for each component $i$ of the vector

- $\theta^i_{GMM} \sim N(\theta_0, T^{-1}((\frac{1}{T}\sum_t \frac{\partial g(\theta^i_{GMM})}{\partial \theta'_{GMM}}' W_T^{i-1}(\frac{1}{T}\sum_t \frac{\partial g(\theta^i_{GMM})}{\partial \theta'_{GMM}})^{-1}$.

- The fully iterative algorithm may be costly to implement if $\theta$ is of large dimension.

Alternative: two step estimator. Start from some $\theta^1_{GMM}$, compute $A, B$, compute $W^1$, compute $\theta^2_{GMM}$ and stop. Properties of two step GMM are the same as those of a fully iterative GMM if initial $\theta^1_{GMM}$ is $T^{0.5}$ consistent.

## 3.1 Complications

• Covariance stationarity is necessary. Linear trends can be accommodated (see Ogaki (1993)) but not unit roots.

• In some economic model the $g$ functions are not martingale difference.

**Example 3.1** *Consider a saving problem with two maturities: one and* $\tau$*. The Euler equations are*

$$E_t[\beta \frac{U_{c,t+1}}{U_{c,t}} - \frac{1}{1+r_{1t}}] = 0 \quad (22)$$

$$E_t[\beta^\tau \frac{U_{c,t+\tau}}{U_{c,t}} - \frac{1}{1+r_{\tau t}}] = 0 \quad (23)$$

$U_{c,t} = \frac{\partial u(c_t)}{\partial c_t}$, $r_{jt}$ *is the real interest rate quoted at* $t$ *for bonds of maturity* $j$. *Then*

$$E_t[\beta^{\tau-1}\frac{U_{c,t+\tau}}{U_{c,t+1}} - \frac{1+r_{1t}}{1+r_{\tau t}}] = 0 \qquad (24)$$

*i.e. expected (at t) forward rate for* $\tau - 1$ *periods must satisfy the no arbitrage condition (24). Log linearizing (24) around the steady state,* $y_{t+\tau} \equiv -\hat{c}_{t+\tau} + \hat{c}_{t+1}$, $x_t \equiv -\frac{r_\tau}{1+r_\tau}\hat{r}_{\tau t} + \frac{r_1}{1+r_1}\hat{r}_{1t}$ *where* $\hat{}$ *represents percentage deviations from the state,* $y_{t+\tau} = \theta x_t + e_{t+\tau}$ *where* $e_{t+\tau}$ *satisfies* $E_t[e_{t+\tau}] = 0$ *and* $\theta_0 = 1$. *If sampling interval of the data different than* $\tau$, $e_{t+\tau}$ *will be serially correlated, e.g. if data is monthly and* $\tau$ *is 12,* $e_{t+\tau}$ *will have MA(11) components.*

What happens if the (sample) orthogonality conditions are not martingale differences?

## A Few Results:

- GMM estimators are consistent and normal in large samples even if $g_t$ is not a martingale difference. However standard errors need to be corrected. In this case, in place of $A$, use $\mathcal{S}(\omega = 0) = \sum_{\tau=-\infty}^{\infty} \frac{1}{T} \sum_t g_t g_{t-\tau} = \sum_{\tau=-\infty}^{\infty} ACF_g(\tau)$.

- Newey-West (1987) Problem: $\mathcal{S}(\omega = 0)$ may not be positive definite. To make sure this is the case need to weight $ACF_g(\tau)$ appropriately i.e. $\mathcal{S}(\omega = 0) = \sum_{\tau=-\infty}^{\infty} \frac{1}{T} \sum_j \mathcal{K}(J(T), \tau) ACF_g(\tau)$. $J(T)$ is truncation point and $\mathcal{K}(J(T), \tau)$ is a kernel.

- Small sample properties of estimators depend on the kernel used.

- All kernel estimates converge very slowly to their true values. Hence, asymptotic approximations are very poor in small samples.

- Approximations may be so bad that it may be preferable to use incorrect standard errors than poorly estimated but correct ones.

$g$ may also be heteroschedastic. If the form of heteroschedasticity is known, get the right expression for A. Otherwise, use a HAC (heteroschedastic consistent covariance) matrix (Newey-West (1987)).

Again, it may be better not to correct for heteroschedasticity if sample is small.

# 4 Testing orthogonality restrictions

a) if $dim(\theta) = dim(z)$ no testing is possible.

b) if $dim(\theta) = q_1 < dim(z) = q$. To estimate $\theta$ need only $q_1$ conditions: $q - q_1$ are left free. If model is correct, also these $q - q_1$ conditions must be close to zero once $\theta_{GMM}$ is plugged in. Hence, under $H_0$.

$$J_T = T \times [g_T(\theta_{GMM})']W_T g_T(\theta_{GMM})] \overset{D}{\sim} \chi^2(q - q_1)$$

where $W_T \overset{P}{\to} A^{-1} = E(g_t(\theta)g_t(\theta)')^{-1}$.

Intuition:

$$y = X\beta + e$$

$z_1, \ z_2$ instruments, $\beta$ is a scalar.

Approximately:

i) get $\hat{\beta}$ using $z_1$;

ii) compute $e(\hat{\beta})$;

iii) calculate $J_T$ using $(z_2 e(\hat{\beta}))^2$. If $T$ is large $J_T$ is $\chi^2(1)$.

# 5 Examples

## 5.1 Estimating linear monetary and fiscal rules

$$
\begin{aligned}
i_t &= b_1\pi_t + b_2 gap_t + b_3 i_{t-1} + u_t && (25)\\
d_t &= a_0 + a_1 d_{t-1} + a_2(DD_t * gap_t) + \\
&+ a_3((1 - DD_t) * gap_t) + a_4 debt_t + e_t && (26)
\end{aligned}
$$

$DD_t =$ dummy variable equal to one if the gap is positive at $t$ (i.e. we are in an expansion) and zero if the gap at $t$ is negative (i.e. we are in a contraction), $d_t$= deficit, $i_t$= nominal interest rate.

Treat $u_t, e_t$ as expectational errors, i.e. $E(u_t|\mathcal{F}_t) = E(e_t|\mathcal{F}_t) = 0$.

Orthogonality conditions (using unconditional expectations):

$$E(i_t - b_1\pi_t - b_2 gap_t - b_3 i_{t-1}) = 0 \quad (27)$$

$$E(d_t - a_0 - a_1 d_{t-1} - a_2(DD_t gap_t) - a_3((1 - DD_t)gap_t) - a_4 debt_t) = 0 \quad (28)$$

Sample counterparts:

$$\frac{1}{T}\sum_t (i_t - b_1\pi_t - b_2 gap_t - b_3 i_{t-1}) = 0 \quad (29)$$

$$\frac{1}{T}\sum_t (d_t - a_0 - a_1 d_{t-1} - a_2(DD_t gap_t) - a_3((1 - DD_t)gap_t) - a_4 debt_t) = 0 \quad (30)$$

Parameters to be estimated $(a_0, a_1, a_2, a_3, a_4, b_1, b_2, b_3)$.

2 equations, 8 unknowns. No more equations in theory. Need to use GIV.

Issues of interest:

- What are the estimates of $b_1, b_2$ and $a_4$?

- Is monetary policy active (i.e. $|b_1| > 1$) and fiscal policy passive $a_4 < 0$ or viceversa? (Leeper (1991))

- Is fiscal policy reacting to output symmetrically over the cycle?

- Do results change with the sample?

Use US data, 1967:1-2004:2. Output gap constructed as actual minus potential output (as reported by the FREDII dataset).

1) Instruments (different for different equations): one lag of output gap, inflation and the nominal interest rate for equation 1; a constant, two lags of deficit, of debt and of inflation and one lag of the interacted variable $DD_t * gap_t$ for equation 2 (total of 11 instruments)

| | b' | se | Ttest |
|---|---|---|---|
| $a_0$ | 0.3653 | 0.1734 | 2.1068 |
| $a_1$ | 0.8393 | 0.0168 | 50.1068 |
| $a_2$ | -0.103 | 0.0128 | -8.0299 |
| $a_3$ | -0.148 | 0.0387 | -3.8295 |
| $a_4$ | -0.0011 | 0.0023 | -0.5009 |
| $b_1$ | 1.3499 | 0.0579 | 23.3134 |
| $b_2$ | 0.0258 | 0.0055 | 4.6816 |
| $b_3$ | 0.7695 | 0.0099 | 78.1224 |

| Jtest |
|---|
| 0.0264 |
| p-value |
| 0.9869 |

| equality test |
|---|
| 2.04E-06 |
| p-value |
| 0.9989 |

Equality test is test of symmetric reaction of deficit to cycle.

2) Use two lags of each instrument in the monetary policy equation. No change in the fiscal equation.

| | b' | se | Ttest |
|---|---|---|---|
| $a_0$ | -0.3236 | 0.1794 | -1.8034 |
| $a_1$ | 0.9031 | 0.0185 | 48.8396 |
| $a_2$ | -0.1217 | 0.0142 | -8.5448 |
| $a_3$ | -0.0096 | 0.0419 | -0.2287 |
| $a_4$ | 0.0078 | 0.0023 | 3.3982 |
| $b_1$ | 0.787 | 0.0413 | 19.0378 |
| $b_2$ | 0.0783 | 0.0039 | 19.9402 |
| $b_3$ | 0.865 | 0.0071 | 122.5776 |

| Jtest |
|---|
| 73.9493 |
| p-value |
| 0 |

| equality test |
|---|
| 1.54E-05 |
| p-value |
| 0.9969 |

3) Omit the 1979-1982 sample.

| | b' | se | Ttest |
|---|---|---|---|
| $a_0$ | 0.3253 | 0.3107 | 1.0472 |
| $a_1$ | 0.8367 | 0.0274 | 30.5522 |
| $a_2$ | -0.129 | 0.0182 | -7.069 |
| $a_3$ | -0.1294 | 0.0629 | -2.0562 |
| $a_4$ | -0.0009 | 0.0041 | -0.2214 |
| $b_1$ | 0.6147 | 0.0663 | 9.2768 |
| $b_2$ | 0.1111 | 0.006 | 18.5329 |
| $b_3$ | 0.891 | 0.0117 | 75.9856 |

| Jtest |
|---|
| 1.6480 |
| p-value |
| 0.4387 |

| equality test |
|---|
| 5.80E-10 |
| p-value |
| 1 |

Here both fiscal and monetary policy are passive $|b_1| < 1$ and $a_4 < 0$ (insignificant). They act as if they have to balance the budget.

4) Eliminate 1979-1982 but use two lags of inflation.

| | b' | se | Ttest |
|---|---|---|---|
| $a_0$ | 0.1859 | 0.3006 | 0.6183 |
| $a_1$ | 0.8439 | 0.0265 | 31.808 |
| $a_2$ | -0.1406 | 0.018 | -7.806 |
| $a_3$ | -0.1037 | 0.0618 | -1.6781 |
| $a_4$ | 0.001 | 0.004 | 0.2439 |
| $b_1$ | 0.5184 | 0.0644 | 8.0496 |
| $b_2$ | 0.1194 | 0.0058 | 20.7261 |
| $b_3$ | 0.9079 | 0.0114 | 79.7907 |

| Jtest |
|---|
| 3.5390 |
| **p-value** |
| 0.1704 |

| equality test |
|---|
| 3.70E-06 |
| **p-value** |
| 0.9985 |

**Estimates unstable. Difficult to draw useful conclusions. Rejection of model may come from first equation (compare 1 and 2).**

How do you check instrument relevance? i.e. that the instruments are correlated with the regressors? Use concentration (F)-statistics.

$$y_t = x_t\beta + e_t$$

$$x_t = z_t\alpha + v_t$$

Construct an F statistics for the hypothesis that $\alpha = 0$. If can't reject, instruments are irrelevant to explain regressors.

Can't use this statistics if the model is nonlinear in the parameters (the distribution of this statistics is unknown in this case).

## 5.2 Estimating a New Keynesian Phillips curve

$$\pi_t = E_t \beta \pi_{t+1} + \frac{(1-\varsigma_p)(1-\varsigma_p \beta)}{\varsigma_p} mc_t \tag{31}$$

where $mc_t = \frac{N_t w_t}{GDP_t}$ are real marginal costs, $\varsigma_p$ is the probability of not changing the prices, $\beta$ the discount factor and $\pi_t$ is the inflation rate.

Transform this equation in an orthogonality condition:

$$E(\pi_t - \beta \pi_{t+1} - \frac{(1-\varsigma_p)(1-\varsigma_p \beta)}{\varsigma_p} mc_t) = 0 \tag{32}$$

Sample counterpart:

$$\frac{1}{T} \sum_t (\pi_t - \beta \pi_{t+1} - \frac{(1-\varsigma_p)(1-\varsigma_p \beta)}{\varsigma_p} mc_t) = 0 \tag{33}$$

Model is a single equation, nonlinear; parameters of interest $(\beta, \varsigma_p)$.

Reduced form orthogonality condition:

$$E_t(\pi_t - a\pi_{t+1} - bmc_t) = 0 \tag{34}$$

Sample counterpart in reduced form linear estimation:

$$\frac{1}{T}\sum_t (\pi_t - a\pi_{t+1} - bmc_t) = 0 \tag{35}$$

Real marginal costs not observable, need a proxy. Could use labor share (LS), or the fact, that in the model, GDP gap (Gap) is proportional to real marginal costs.

Use a constant and up to 5 lags of inflation and marginal costs as instruments. Report the result most favorable to the theory

**Table: Estimates of NK Philips curve**

| | Linear estimates | | Nonlinear estimates | | |
|---|---|---|---|---|---|
| Country/Proxy | $a$ | $b$ | $\beta$ | $\varsigma_p$ | J-Test p-value |
| US-Gap | 0.993(16.83) | -0.04 (-0.31) | 0.907 (10.35) | 0.700 (5.08) | $\chi^2(5)$=0.15 |
| US-LS | 0.867(10.85) | 0.001( 1.75) | 0.932 ( 7.74) | 0.991 (150.2) | $\chi^2(9)$=0.54 |
| UK-Gap | 0.667(7.18) | 0.528( 1.30) | 0.924 ( 4.96) | 0.684 (1.37) | NA |
| UK-LS | 0.412(3.81) | 0.004( 4.05) | 0.853 ( 4.07) | 0.994 (166.1) | $\chi^2(1)$=0.25 |
| GE-Gap | 0.765(10.02) | -0.01 (-0.22) | 0.972 (7.48) | 1.014 (0.03) | NA |
| GE-LS | 0.491(3.34) | 0.03 ( 1.85) | 0.919 ( 4.45) | 0.958 (7.18) | $\chi^2(1)$=0.83 |

## 5.3 Estimating a RBC model

Model has multiple equations and is nonlinear.

$\max_{(c_t, K_{t+1}, N_t)} E_0 \sum_t \beta^t \frac{c_t^{1-\varphi}}{1-\varphi} + \vartheta_N (1 - N_t)$ subject to

$$g_t + c_t + K_{t+1} = \zeta_t K_t^{1-\eta} N_t^{\eta} + (1-\delta)k_t = GDP_t + (1-\delta)K_t \qquad (36)$$

$\ln \zeta_t = \bar{\zeta} + \rho_z \ln \zeta_{t-1} + \epsilon_{1t} \quad \epsilon_{1t} \sim (0, \sigma_z^2)$,
$\ln g_t = \bar{g} + \rho_g \ln g_{t-1} + \epsilon_{2t} \quad \epsilon_{2t} \sim (0, \sigma_g^2)$, $k_0$ given.

Let $g_t$ be financed with lump sum taxes. Optimality conditions:

$$\vartheta_N c_t^{\varphi} = \eta \zeta_t K_t^{1-\eta} N_t^{\eta-1} \qquad (37)$$

$$1 = E_t \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\varphi} [(1-\eta)\zeta_{t+1} K_{t+1}^{-\eta} N_{t+1}^{\eta} + (1-\delta)] \qquad (38)$$

and $w_t = \eta \frac{GDP_t}{N_t}$, $r_t = (1-\eta)\frac{GDP_t}{K_t} + (1-\delta)$.

11 parameters: five structural $\alpha_1 = (\beta, \vartheta_N, \varphi, \eta, \delta)$ and 6 auxiliary ones $\alpha_2 = (\bar{\zeta}, \bar{g}, \rho_z, \rho_g, \sigma_g, \sigma_z)$. Need at least 11 orthogonality conditions.

$$E(\delta - 1 + \frac{k_{t+1}}{k_t} - \frac{inv_t}{k_t}) = 0 \tag{39}$$

which determines $\delta$ if capital and investment data are available.

$$E_t\beta(\frac{c_{t+1}}{c_t})^{-\varphi})[(1-\eta)\zeta_{t+1}K_{t+1}^{-\eta}N_{t+1}^{\eta} + (1-\delta)] = 0 \tag{40}$$

contains four parameters $(\beta, \varphi, \eta, \delta)$. Since $\delta$ is "identified" from (39) transform (40) to produce at least three orthogonality conditions. Given

$\delta$, the three parameters could be estimated using

$$
\begin{aligned}
E(\beta(\frac{c_{t+1}}{c_t})^{-\varphi})[(1-\eta)\zeta_{t+1}K_{t+1}^{-\eta}N_{t+1}^{\eta}+(1-\delta)]) &= 0 \quad (41)\\
E(\beta(\frac{c_{t+1}}{c_t})^{-\varphi})[(1-\eta)\zeta_{t+1}K_{t+1}^{-\eta}N_{t+1}^{\eta}+(1-\delta)]\frac{c_t}{c_{t-1}}) &= 0 \quad (42)\\
E(\beta(\frac{c_{t+1}}{c_t})^{-\varphi})[(1-\eta)\zeta_{t+1}K_{t+1}^{-\eta}N_{t+1}^{\eta}+(1-\delta)]r_t) &= 0 \quad (43)
\end{aligned}
$$

The intratemporal condition

$$
E[c_t^{-\varphi}\eta\zeta_t K_t^{1-\eta}N_t^{\eta-1}-\vartheta_N]=0 \quad (44)
$$

involves $(\varphi, \eta, \vartheta_N)$. Given $(\varphi, \eta)$, it determines $\vartheta_N$.

The auxiliary parameters can be estimated using

$$
\begin{aligned}
E(\ln \zeta_t - \bar{\zeta} + \rho_z \ln \zeta_{t-1}) &= 0 && (45)\\
E(\ln \zeta_t - \bar{\zeta} + \rho_z \ln \zeta_{t-1}) \ln \zeta_{t-1} &= 0 && (46)\\
E(\ln \zeta_t - \bar{\zeta} + \rho_z \ln \zeta_{t-1}))^2 - \sigma_z^2 &= 0 && (47)\\
E(\ln g_t - \bar{g} + \rho_g \ln g_{t-1}) &= 0 && (48)\\
E(\ln g_t - \bar{g} + \rho_g \ln g_{t-1}) \ln g_{t-1} &= 0 && (49)\\
E(\ln g_t - \bar{g} + \rho_g \ln g_{t-1}))^2 - \sigma_g^2 &= 0 && (50)
\end{aligned}
$$

• Government expenditure is observable, technological disturbances are not. One additional auxiliary condition needed. From production function, given estimates of $\eta$; $\hat{z}_t = \ln GDP_t - (1-\eta)\ln K_t - \eta \ln N_t$.

• Sequential estimation: the last three conditions estimable separately from first eight. (Need to correct for s.e. if you do this). Otherwise go joint.

- Here use just identified system or a weakly overidentified one, without optimal weighting matrix . Overidentified estimates obtained adding lags of $\frac{GDP_t}{K_t}, \frac{c_t}{c_{t-1}}$, of investment or of the output labor ratio.

Linearly detrended US quarterly data; 1956:1-1984:1.

**Table: Estimates of a RBC model**

| Parameter | just identified | over identified | fixing $\eta$ | $\varphi = 2$ |
|---|---|---|---|---|
| $\eta$ | 0.18 (0.0002) | 0.18 (0.0002) | 0.66 | 0.18 (0.0002) |
| $\varphi$ | 1.0 | 1.0 | 1.0 | 2.0 |
| $\delta$ | 0.015(0.043) | 0.019 (0.042) | 0.017 (0.038) | 0.04 (0.039) |
| $\beta$ | 0.980 (0.018) | 0.955 (0.033) | 0.940 (0.021) | 0.830 (0.068) |
| $\vartheta_N$ | 81.39 (65.43) | 82.11 (72.32) | 109.07 (148.21) | 4.42 (1.987) |
| $\bar{\zeta}$ | 0.0005 (0.001) | -0.0002 (0.001) | 0.001(0.001) | 0.0008 (0.001) |
| $\bar{g}$ | 0.075(0.001) | 0.076(0.001) | 0.076 (0.001) | 0.076 (0.001) |
| $\rho_z$ | 1.038(0.054) | 1.023 (0.050) | 1.032 (0.050) | 1.008 (0.049) |
| $\rho_g$ | 0.998 (0.0005) | 0.998 (0.0006) | 0.998 (0.0006) | 0.998 (0.0006) |
| $\sigma_z^2$ | 0.0001 (0.00001) | 0.0001 (0.00001) | 0.0001 (0.00001) | 0.0001 (0.00001) |
| $\sigma_z^2$ | 0.0002 (0.00002) | 0.0002 (0.00002) | 0.0002 (0.00002) | 0.0002 (0.00002) |
| $\chi^2(6)$ | | 378.00 | 345.31 | 370.55 |

- Conditional on $\varphi$, very low estimate of $\eta$ is obtained.

- Structural parameters imprecisely estimated (except for $\beta$).

- AR parameters in the near non-stationary region (both with just-identified or overidentified systems).

- Estimates of $\beta$ are economically unreasonable except in the just-identified system.

- Model strongly rejected in all cases ($\chi^2$ statistic smaller in last two cases).

- Results are broadly independent of the value of $\varphi$ chosen.

## Economic analyses with GMM estimates

- Tests of orthogonality conditions often not very informative about why the model is rejected.

- What does rejection of orthogonality conditions imply in economic terms?

Let $h(\theta)$ be some statistics you can construct from the model (plug in estimate and solve it) and $h_T$ be the statistics you can construct in the data (with sample size $T$).

h=mean, variances, autocorrelations, etc.

Let $m(\theta_T) = h(\theta_T) - h_T$. Then $\Sigma_T = (\frac{\partial h(\theta_0)}{\partial \theta'})(\Sigma_\theta)(\frac{\partial h(\theta_0)}{\partial \theta'}) + \Sigma_h$.

Under $H_0$: $Tm(\theta_T)'\Sigma_T^{-1}m(\theta_T) \xrightarrow{D} \chi^2(dim(h))$.

• This test can be done for any subsets of statistics of interest.

• To construct $h(\theta)$ need to solve and simulate model (GMM estimation does not require solving the model, just uses the FOCs).

Using just-identified estimates (after log-linearization).

**Economic Tests of the RBC model**

| Moment | P-value | Moment | P-value | Moment | P-value |
|---|---|---|---|---|---|
| var(c)/var(gdp) | 0.02 | c-AR(1) | 0.03 | corr(c,gdp) | 0.09 |
| var(inv)/var(gdp) | 0.00 | inv-AR(1) | 0.00 | corr(inv,gdp) | 0.01 |
| var(N)/var(gdp) | 0.00 | N-AR(1) | 0.00 | corr(N,gdp) | 0.00 |
| | | gdp-AR(1) | 0.00 | | |

# 6 Relationship GMM/Calibration

- In GMM (just identified case) find $\theta$ so that $\frac{1}{T}\sum_t g_t(\theta) = 0$.

- In calibration find $\theta$ so that $\frac{1}{T}\sum_t y_t(\theta) = \bar{y}$ ($\bar{y}$ are the steady states).

If we set $g_t = y_t - \bar{y}$

• Calibration is equivalent to a just-identified GMM when the orthogonality condition is the deviation of the variable from the steady state.

If we add conditions like $var(c_t(\theta)) = var(c_t)$, then $g_t = [g_{1t}, g_{2t}]$, $g_{1t} = y_t - \bar{y}$ and $g_{2t} = var(c_t(\theta)) - var(c_t)$.

• Calibration to the steady state with additional conditions is equivalent to just-identified GMM on $[g_{1t}, g_{2t}]$.

# 7 General problems with GMM:

- Choice of instruments: which one? How many ?

- Choice of orthogonality conditions ? Which one do we select?

- Small sample approximation? Generally bad.

- What should we do to correct for deviation of data from ideal conditions?

- Identification of parameters? $B$ may be close to or exactly singular.

- What if there are unobservable variables in orthogonality conditions?

# 8 Exercises

1) Consider estimating a log-linearized Euler equation with GIV when consumer's utility displays external habit persistence in consumption. In this case the Euler equation can be written as:

$$\frac{\gamma}{1+\gamma}\Delta c_t = \frac{1}{1+\gamma}E_t\Delta c_{t+1} + \frac{1-\gamma}{(1+\gamma)\phi}(i_t - E_t\pi_{t+1}) \tag{51}$$

where $\Delta c_t$ is consumption growth, $i_t$ the nominal interest rate and $\pi_t$ inflation. There are two parameters: the risk aversion coefficient $\phi$ and the habit parameter $\gamma$. Using both a just-identified and an overidentified system, test the relevance of consumption habit in matching the data of the country you are considering. (Hint: you will have to make a lot choices here: which instruments to use, which consumption series to use, how to deal with potential non-stationarities of the inflation rate, etc. Make sure you are very transparent in specifying what you do. Most of the grade in this question depends on how well you explain what you are doing).

2) Consider three monetary policy rules of the form

$$\begin{aligned} i_t &= \rho i_{t-1} + (1-\rho)\phi_x x_t + (1-\rho)\phi_p \pi_t + e_t && (52) \\ i_t &= \rho i_{t-1} + (1-\rho)\phi_x x_{t-1} + (1-\rho)\phi_p \pi_{t-1} + e_t && (53) \\ i_t &= \rho i_{t-1} + (1-\rho)\phi_x E_t x_{t+1} + (1-\rho)\phi_p E_t \pi_{t+1} + e_t && (54) \end{aligned}$$

The difference between various specifications is that the central bank looks at the past, the present or the future output gap ($x_t$) and inflation rates ($\pi_t$) when setting interest rates. Estimate the parameters of these three rules by GIV using data for your favorite country. Which specification fits the data better? Why? Test the first specification against the third (Hint: estimate a general model and test the three models as restricted specifications).

3) Consider an extended Phillips curve equation of the form:

$$\pi_t = \frac{\beta}{1+\beta\gamma_p} E_t \pi_{t+1} + \frac{\gamma_p}{1+\beta\gamma_p} \pi_{t-1} + \frac{(1-\varsigma_p)(1-\varsigma_p\beta)}{\varsigma_p} mc_t + e_t \qquad (55)$$

where $mc_t = \frac{N_t w_t}{GDP_t}$ are real marginal costs, $\varsigma_p$ is the probability of not changing the prices, $\pi_t$ is the inflation rate and $\gamma_p$ is an indexation parameter.

Estimate the parameters of this model by GIV. Test whether indexation is important. Repeat estimation fixing $\beta = 0.99$ (assuming you have quarterly data).