A Labour Econometric Example for GMM

Consider the case were the interest lies in the determination of the conditional expectation of wage, as an exponential function of hours worked and education. It is thus assumed that for any individual i, her wage y_i, has a conditional, on the hours worked x_{1i}, and the education years x_{2i} of the form $\mathbb{E}(y_i / x_{1i} x_{2i}) =$
$= \exp(\theta_{10} x_{1i} + \theta_{20} x_{2i})$. A sample generated by n individuals is available,

of the form (Y_n, X_n) where

$$Y_n = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X_n = \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ \vdots & \\ x_{1n} & x_{2n} \end{pmatrix},$$

the random elements $(y_i, x_{1i}, x_{2i})_{i=1,\dots,n}$ are independent, and given that the vector $\theta_0 = \begin{pmatrix} \theta_{10} \\ \theta_{20} \end{pmatrix}$ is latent, the statistical/econometric problem is the approximation of θ_0, from the information present in the sample and the statistical model to be determined below: it is assumed that $\begin{pmatrix} x_{11} \\ \vdots \\ x_{1n} \end{pmatrix}$ is observed with measurement error which implies that in the relation $y_i = \exp(X_{(i)}\theta_0) + \varepsilon_i$

$\mathbb{E}(\varepsilon_i \mid x_{1i}) \neq 0$ at least for some i.

This would result into inconsistency for an estimator as NLLSE.

It is also assumed that a sample of instruments are observed, without measurement error, $W_n = \begin{pmatrix} w_{11} & w_{21} & x_{21} \\ w_{12} & w_{22} & x_{22} \\ \vdots & \vdots & \vdots \\ w_{1n} & w_{2n} & x_{in} \end{pmatrix}$ where

w_{1i} is the family size of individual i, and W_{2i} is the non labour income of this individual* The analyst reasonably expects that $\mathbb{E}(\varepsilon_i / W_{(i)}) = 0 \ \forall i=1,\dots,n$. Hence

See employs the following IV linear exponential semi-parametric (why?)

* $(\varepsilon_i, x_{1i}, x_{2i}, w_{1i}, w_{2i})_{i=1,\dots,n}$ are also assumed to have iid rows

Model $Y_n = EXP(X_n\theta) + \varepsilon_n$

and $\mathbb{E}(\varepsilon_n \mid \sigma(W_n)) = 0_{n\times 1}$

and $\theta \in \Theta$, which is assumed without much loss of of generality to be a non-empty compact subset of $\mathbb{R}^3$, and $\theta_0 \in \Theta$; also remember that

$$EXP(X) := \begin{pmatrix} \exp(x_1) \\ \exp(x_2) \\ \vdots \\ \exp(x_n) \end{pmatrix}.$$

Given this structure, the 2-GMME is

Considered as:

$$\hat{\theta}_n := \underset{\theta\in\mathbb{R}^3}{\arg\min} \left(Y_n' - EXP'(X_n\theta)\right) W_n W_n' \left(Y_n - EXP(X_n\theta)\right)$$

$$\theta_n\left(V_n^{-1}(\hat{\theta}_n)\right) =$$

$$\underset{\theta\in\mathbb{R}^3}{\arg\min} \left(Y_n' - EXP'(X_n\theta)\right) W_n V_n^{-1}(\hat{\theta}_n) W_n' \left(Y_n - EXP(X_n\theta)\right)$$

where (due to independence)

$V_n(\hat{\Theta}_n) :=$

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_{(i)} - \exp(x_{(i)}\hat{\Theta}_n)\right) W_{(i)} W'_{(i)} \left(y_{(i)} - \exp(x_{(i)}\hat{\Theta}_n)\right)$$

where $W_{(i)} = (w_{1i} \; w_{2i} \; x_{2i})$.

Regarding the asymptotic properties of the estimator in this particular case:

- $m(z,\theta) := w(y - \exp(x\theta))$ which is continuous in θ w.r.t. $z = (y, x, w)$

- $\mathbb{E}(\| w'_{(1)} (y_1 - \exp(x_1\theta)) \|) =$

$$= \mathbb{E}\left[\left\| \begin{pmatrix} w_{11} \\ w_{21} \\ w_{31} \end{pmatrix} \left(y_1 - \exp(x_{11}\theta_1 + x_{12}\theta_2)\right) \right\|\right] =$$

$$= \mathbb{E}\left[\sum_{i=1}^{3}\left(w_{i1}\left(y_1 - \exp(x_{11}\theta_1 + x_{12}\theta_2)\right)\right)^2\right]^{1/2}, \text{ and}$$

this is less than or equal to

[$(\sum|a_i|)^2 \geq \sum|a_i|^2$, $x \to \sqrt{x}$ and $\mathbb{E}(\cdot)$ are Monotone]

$$(*) \quad \sum_{i=1}^{3}\mathbb{E}\left[\left|w_{i1}\left(y_1 - \exp(x_{11}\theta_1 + x_{12}\theta_2)\right)\right|\right]$$

and since $y_1 = \exp(x_{11}\theta_{10} + x_{12}\theta_{20}) + \varepsilon_1$

(*) is less than or equal to [$|\alpha+b| \leq |\alpha| + |b|$]

$$\sum_{i=1}^{3}\mathbb{E}\Big[|w_{i1}\varepsilon_1| + |w_{i1}\exp(x_{11}\theta_1 + x_{12}\theta_2)|$$

$$+ |w_{i1}\exp(x_{11}\theta_{10} + x_{12}\theta_{20})|\Big]$$

$$= \sum_{i=1}^{3}\mathbb{E}\left[|w_{i1}\varepsilon_1|\right] + \sum_{i=1}^{3}|w_{i1}\exp(x_{11}\theta_1 + x_{12}\theta_2)|$$

$$+ \sum_{i=1}^{3}\mathbb{E}\left[|w_{i1}\exp(x_{11}\theta_{10} + x_{12}\theta_{20}|\right]$$

Hence if $\max_{i=1,\dots,3}\mathbb{E}[|w_{i1}\varepsilon_1|] + \max_{i=1,\dots,3}\mathbb{E}[|w_{i1}\exp(x_{11}\theta_1 + x_{12}\theta_2)|]$

$< +\infty \quad \forall \theta \in \Theta$ (d), due to the iidness and the standard LLN,

$$\frac{1}{n}\sum_{i=1}^{n} W_{(i)}'(Y_i - \exp(X_{(i)}\theta)) \xrightarrow{P}$$

$$\mathbb{E}[W_{(1)}(Y_1 - \exp(X_{(1)}\theta))] \quad \forall \theta \in \Theta$$

Notice also that if $\theta^*, \theta_* \in \Theta$

$$\| W_{(i)}'(Y_i - \exp(X_{(i)}\theta^*)) - W_{(i)}'(Y_i \exp(X_{(i)}\theta_*))\|$$

$$= \| W_{(i)}'[\underbrace{\exp(X_{(i)}\theta_*) - \exp(X_{(i)}\theta^*)}_{\text{scalar}}]\| =$$

$$= \|W_{(i)}\| \, |\exp(X_{(i)}\theta_*) - \exp(X_{(i)}\theta^*)|$$

MVT

$$\leq \|W_{(i)}\| \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta'} \exp(X_{(i)}\theta) \right\| \, \|\theta_* - \theta^*\|$$

$$= \|W_{(i)}\| \sup_{\theta \in \Theta} \| X_{(i)} \underbrace{\exp(X_{(i)}\theta)}_{\text{scalar}} \| \, \|\theta_* - \theta^*\|$$

$$= \sup_{\theta \in \Theta} \exp(X_{(i)}\theta) \, \|W_{(i)}\| \, \|X_{(i)}\| \, \|\theta_* - \theta^*\|,$$

which then means that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} W_{(i)} (Y_i - \exp(X_{(i)} \theta^*)) - \frac{1}{n} \sum_{i=1}^{n} W_{(i)} (Y_i - \exp(X_{(i)} \theta_*)) \right\|$$

tr. ineq.

$$\leq \frac{1}{n} \sum_{i=1}^{n} \| W_{(i)} (Y_i - \exp(X_{(i)} \theta^*)) - W_{(i)} (Y_i - \exp(X_{(i)} \theta_*)) \|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left[\sup_{\theta \in \Theta} \exp(X_{(i)} \theta) \| W_{(i)} \| \, \| X_{(i)} \| \right] \| \theta^* - \theta_* \|$$

And if $\mathbb{E}\left[\sup_{\theta \in \Theta} \exp(X_{(1)} \theta) \| W_{(1)} \| \, \| X_{(1)} \| \right] < +\infty$, (b)

then due to the classical LLN

$$\frac{1}{n} \sum_{i=1}^{n} \left[\sup_{\theta \in \Theta} \exp(X_{(i)} \theta) \| W_{(i)} \| \, \| X_{(i)} \| \right] = O_p(1).$$

Combining the previous it is obtained that

if (a) and (b) hold

$$\frac{1}{n}\sum_{i=1}^{n} W_{(i)}(y_i - \exp(X_{(i)}\theta)) \xrightarrow{P} \mathbb{E}[W_{(1)}(y_1 - \exp(X_{(1)}\theta))]$$

Furthermore suppose that (c)

$$\mathbb{E}[W_{(1)}[\exp(X_{(1)}\theta) - \exp(X_{(1)}\theta_0)]] = 0_{3\times 1}$$

iff $\theta = \theta_0$, which is absolutely plausible since all the random variables employed are non-negative [both in W_n and in X_n], and the exponential is an injective function.

Given that by assumption $\mathbb{E}(W_{(1)}\varepsilon_1) = 0_{3\times 1}$ and $y_{(i)} = \exp(X_{(i)}\theta_0) + \varepsilon_i$, (c) implies the identification condition in our main consistency theorem. Hence the latter implies that under (a), (b), (c), $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Now consider

$$V_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} (y_i - \exp(x_{(i)}\theta))' w_{(i)} w_{(i)}' (y_i - \exp(x_{(i)}\theta))$$

$$= \frac{1}{n}\sum_{i=1}^{n} [\varepsilon_i + \exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)]^2 w_{(i)} w_{(i)}'$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 w_{(i)} w_{(i)}'}_{(A(\theta))} + 2\underbrace{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i (\exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)) w_{(i)} w_{(i)}'}_{(B(\theta))}$$

$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n} [\exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)]^2 w_{(i)} w_{(i)}'}_{(T(\theta))}.$$

Using an analogous analysis it is possible to show that if $\max_{i=1,\dots,3} \mathbb{E}(\varepsilon_i^2 w_i^2) < +\infty$ then (d)

$\mathbb{E}[\varepsilon_i^2 w_{(i)} w_{(i)}']$ exists and due to the standard LLN $\underbrace{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 w_{(i)} w_{(i)}'}_{(A'(\theta))} \xrightarrow{P} \mathbb{E}(\varepsilon_i^2 w_{(i)} w_{(i)}')$.

Furthermore, if

$$\max_{i=1,\dots,3} \mathbb{E}\left[|\varepsilon_1| \sup_{\theta\in\Theta} \exp(x_{(1)}\theta)\, w_{i1}^2\right] < +\infty \text{ then } (e)$$

$\mathbb{E}\left[\varepsilon_1 \left(\exp(x_{(1)}\theta_0) - \exp(x_{(1)}\theta)\right) w_{(1)} w_{(1)}'\right]$ exists

and $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \left(\exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)\right) w_{(i)} w_{(i)}'$

$$\xrightarrow{cp} \underbrace{\mathbb{E}\left[\varepsilon_1 \left(\exp(x_{(1)}\theta_0) - \exp(x_{(1)}\theta)\right) w_{(1)} w_{(1)}'\right]}_{(B'(\theta))}$$

And Analogously if

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \left(\exp(2x_{(1)}\theta)\right) w_{(1)} w_{(1)}'\right] < +\infty \text{ then } (f)$$

$\mathbb{E}\left[\left(\exp(x_{(1)}\theta_0) - \exp(x_{(1)}\theta)\right)^2 w_{(1)} w_{(1)}'\right]$ exists and

$\frac{1}{n}\sum_{i=1}^{n} \left(\exp(x_{(i)}\theta_0) - \exp(x_{(i)}\theta)\right)^2 w_{(i)} w_{(i)}'$

$$\xrightarrow{cp} \underbrace{\mathbb{E}\left[\left(\exp(x_{(1)}\theta_0) - \exp(x_{(1)}\theta)\right)^2 w_{(1)} w_{(1)}'\right]}_{(\Gamma'(\theta))}$$

Since $\forall \theta^* \to \theta$, $\forall \varepsilon > 0$

$$\mathbb{P}(|A(\theta^*) - A'(\theta)| > \varepsilon) \to 0$$

$$\mathbb{P}(|B(\theta^*) - B'(\theta)| > \varepsilon) \to 0$$

$$\mathbb{P}(|\Gamma(\theta^*) - \Gamma'(\theta)| > \varepsilon) \to 0$$

we also have that $\forall \theta^* \to \theta$, $\forall \delta > 0$

$$\mathbb{P}\left[|A(\theta^*) + B(\theta^*) + \Gamma(\theta^*) - (A'(\theta) + B'(\theta) + \Gamma'(\theta))| > \delta\right]$$

$$= \mathbb{P}\left[|(A(\theta^*) - A'(\theta)) + (B(\theta^*) - B'(\theta)) + (\Gamma(\theta^*) - \Gamma'(\theta))| > \delta\right]$$

tri.ineq.

$$\leq \mathbb{P}\left[|A(\theta^*) - A'(\theta)| + |B(\theta^*) - B(\theta)| + |\Gamma(\theta^*) - \Gamma(\theta)| > \delta\right]$$

$$\leq \mathbb{P}\left[|A(\theta^*) - A'(\theta)| > \delta/3\right] + \mathbb{P}\left[|B(\theta^*) - B'(\theta)| > \delta/3\right]$$

$$+ \mathbb{P}\left[|\Gamma(\theta^*) - \Gamma'(\theta)| > \delta/3\right] \to 0,$$

establishing finally that

$V_n(\theta) \xrightarrow{cp} V(\theta)$ with

$V(\theta) := \Lambda'(\theta) + B'(\theta) + \Gamma'(\theta).$

Using the above and the fact that $V_n(\theta)$ and $\hat{\theta_n}$ become asymptotically independent it is then possible to prove that:

$$V_n(\hat{\theta_n}) \xrightarrow{P} V(\theta_0) = \Lambda'(\theta_0) + B'(\theta_0) + \Gamma'(\theta_0)$$

with $\Lambda'(\theta_0) = \mathbb{E}\left[\varepsilon_L^2 W_{(L)} W_{(1)}'\right]$

$$B'(\theta_0) = \mathbb{E}\left[\varepsilon_L\left(\exp(X_{(1)}\theta_0) - \exp(X_{(1)}\theta_0)\right) W_{(1)} W_{(1)}'\right]$$

$$= \mathbb{E}\left(\varepsilon_L \cdot 0 \cdot W_{(1)} W_{(1)}'\right) = 0_{3\times 3}$$

and $\Gamma'(\theta_0) = \mathbb{E}[(\exp(x_{(1)}\theta_0) - \exp(x_{(1)}\theta_0))^2 W_{(1)} W_{(1)}']$

$= \mathbb{E}[0\, W_{(1)} W_{(1)}'] = 0_{3\times 3}.$

Hence $\sqrt{n}(\hat{\theta}_n) \to \mathbb{E}[\varepsilon_i^2 W_{(1)} W_{(1)}']$.

If moreover $\operatorname{rank}(W_n W_n') = 3$ and ε_i is non degenerate at zero (g), then $\mathbb{E}(\varepsilon_i^2 W_{(1)} W_{(1)}')$ is non singular, and thereby our general consistency theorem applies to the 2-GMME establishing that under (a) - (g), $\theta_n \xrightarrow{P} \theta_0$.

Regarding the rate and the limiting distribution notice that:

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{(i)}\varepsilon_i \rightsquigarrow N\left(0_{3\times1}, \mathbb{E}\left[\varepsilon_i^2 W_{(i)} W_{(i)}'\right]\right)$$

due to the iid-ness framework, (d), and the classical CLT.

$$-\frac{\partial W_{(i)}'(y_i - \exp(X_{(i)}\theta))}{\partial\theta'} =$$

$$= -\overbrace{\underbrace{W_{(i)}'}_{3\times1}\underbrace{X_{(i)}}_{1\times2}}^{3\times2}\underbrace{\exp(X_{(i)}\theta)}_{1\times1}$$

and thereby $\frac{1}{n}\sum_{i=1}^{n}\frac{\partial W_{(i)}'(y_i - \exp(X_{(i)}\theta))}{\partial\theta'}$

$$= -\frac{1}{n}\sum_{i=1}^{n} W_{(i)}' X_{(i)} \exp(X_{(i)}\theta)$$

which if (B) holds it converges continuously in probability to $-\mathbb{E}[W'_{(i)} X_{(i)} \exp(X_{(i)}\theta)]$,

$$-\frac{\partial^2 W'_{(i)}(y_i - \exp(X_{(i)}\theta))}{\partial\theta'\partial\theta_j} = W'_{(i)} X_{(i)} X_{(i,j)} \exp(X_{(i)}\theta), \quad \forall j = 1,2$$

and thereby $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 W_{(i)}(y_i - \exp(X_{(i)}\theta))}{\partial\theta'\partial\theta_j}$

$= \frac{1}{n}\sum_{i=1}^{n} W'_{(i)} X_{(i)} X_{(i,j)} \exp(X_{(i)}\theta)$ and if (B) is enforced to

$$\mathbb{E}\left[\sup_{\theta} \exp(X_{(i)}\theta) \|W_{(i)}\| \|X_{(i)}\|\right] < +\infty \quad (B^*),$$ it

can be shown to be $O_p(1)$ uniformly on Θ.

Thereby, under (a), (b*), (c) – (f) our theory finally says that for the 2-GMME

$$n^{1/2}(\theta_n - \theta_0) \rightsquigarrow N(0_{2\times 1}, V_*) \text{, with}$$

$$V_* := \mathbb{E}[\underbrace{x_{(i)}'}_{2\times 1} \underbrace{w_{(i)}}_{1\times 3} \underbrace{\exp(x_{(i)}\theta_0)}_{1\times 1}] \, \mathbb{E}\underbrace{(\varepsilon_i^2 w_{(i)} w_{(i)}')}_{3\times 3}^{-1} \mathbb{E}[w_{(i)}' x_{(i)} \exp(x_{(i)}\theta_0)]$$

Due to the previous a consistent estimator for $\mathbb{E}[w_{(i)}' x_{(i)} \exp(x_{(i)}\theta_0)]$ is

$$\frac{1}{n}\sum_{i=1}^{n} w_{(i)}' x_{(i)} \exp(x_{(i)}\theta_n) \quad \text{(why?)}$$

while for $e_i := y_i - \exp(x_{(i)}\theta_n)$

$$\frac{1}{n}\sum_{i=1}^{n} (y_i - \exp(x_{(i)}\theta_n))^2 w_{(i)} w_{(i)}' =$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i + \exp(X_{(i)}\theta_0) - \exp(X_{(i)}\theta_n)\right)^2 W_{(i)} W_{(i)}'$$

which we know that it converges in probability to $\mathbb{E}[\varepsilon_i^2 W_{(i)} W_{(i)}']$ from the above, and thereby due to the non-singularity of $\mathbb{E}(\varepsilon_i^2 W_{(i)} W_{(i)}')$ and the CMT,

$$\left[\frac{1}{n}\sum_{i=1}^{n} e_i^2 W_{(i)} W_{(i)}'\right]^{-1} \xrightarrow{P} \mathbb{E}[\varepsilon_i^2 W_{(i)} W_{(i)}']^{-1}$$

$\downarrow$

And this is at least asymptotically well defined.

Hence the CMT then implies that

$$V_n^* := \frac{1}{n}\left[\sum_{i=1}^{n} X_{(i)}' W_{(i)} \exp(X_{(i)}\theta_n)\left(\sum_{i=1}^{n} e_i^2 W_{(i)} W_{(i)}'\right)^{-1} \times \sum_{i=1}^{n} W_{(i)}' X_{(i)} \exp(X_{(i)}\theta_n)\right] \text{ is}$$

a consistent estimator of V^*.

Note The conditions on the existence of moments that appear in this text would directly hold if w_n, and x_n were comprised by bounded random variables - something that is not completely implausible given the economic nature of those variables - and $\mathbb{E}(\varepsilon_i^2) < +\infty$.

Note The compactness of Θ can be dispensed. The conditions involving $\sup_{\theta \in \Theta}$ can be weakened to hold locally uniformly.

Notice also that the statistic

J_n can be used here in order to test

the hypothesis that $\mathbb{E}(W'_{(n)}\varepsilon_{(n)}) = 0_{3\times1}$,

given the identification condition (C).

Under this null $J_n \rightsquigarrow \chi^2_1$ which is

usable for inference.

Matlab Code for the simulation, GMM estimation and J-test inference on this model.

```
% Non-Linear GMM Estimation with J-Test in MATLAB
clear; clc; rng(42);

% Simulated Data
N = 500; % Sample size
family_size = randi([1, 5], N, 1); % Instrument 1
non_labour_income = normrnd(20000, 5000, N, 1); % Instrument 2
education = randi([8, 20], N, 1); % Instrument 3
Z = [family_size, non_labour_income, education]; % Instrument matrix

% True Parameters
theta_true = [0.05, 0.1]; % [theta1, theta2]

% Generate Endogenous Variable (hours worked)
hours = 40 + 2 * family_size - 0.001 * non_labour_income + 0.5 * randn(N, 1);

% Generate Wages (dependent variable)
epsilon = normrnd(0, 100, N, 1);
wage = exp(theta_true(1) * hours + theta_true(2) * education) + epsilon;

% Define the moment conditions
function g = gmm_moments(theta, wage, hours, education, Z)
   % Residual
   residual = wage - exp(theta(1) * hours + theta(2) * education);
   % Moment conditions
   g = (Z' * residual) / length(wage);
end

% Define the GMM objective function
function Q = gmm_objective(theta, wage, hours, education, Z, W)
   % Compute moments
   g = gmm_moments(theta, wage, hours, education, Z);
   % Compute GMM objective
   Q = g' * W * g;
end

% Initial parameter guesses
theta_init = [0.01, 0.01];

% Step 1: Use identity weighting matrix
W_identity = eye(size(Z, 2));
options = optimoptions('fminunc', 'Display', 'iter', 'Algorithm', 'quasi-newton');
[theta_step1, Q_step1] = fminunc(@(theta) gmm_objective(theta, wage, hours, education, Z, W_identity), theta_init, options);

% Step 2: Update weighting matrix using residuals
g_step1 = gmm_moments(theta_step1, wage, hours, education, Z);
S_hat = cov(g_step1'); % Variance-covariance of moment conditions
W_optimal = inv(S_hat); % Optimal weighting matrix

% Step 3: Re-estimate using optimal weighting matrix
[theta_step2, Q_step2] = fminunc(@(theta) gmm_objective(theta, wage, hours, education, Z, W_optimal), theta_step1, options);

% Compute the J-statistic
g_step2 = gmm_moments(theta_step2, wage, hours, education, Z);
J_stat = N * (g_step2' * W_optimal * g_step2); % J-statistic
df = size(Z, 2) - length(theta_step2); % Degrees of freedom
p_value = 1 - chi2cdf(J_stat, df);

% Display results
fprintf('Estimated Parameters (Two-Step GMM):\n');
fprintf('Theta_1: %.4f, Theta_2: %.4f\n', theta_step2(1), theta_step2(2));
fprintf('\nJ-Test for Overidentifying Restrictions:\n');
fprintf('J-statistic: %.4f\n', J_stat);
fprintf('Degrees of Freedom: %d\n', df);
fprintf('P-value: %.4f\n', p_value);
```

Python code for the simulation, the GMM estimation, and the J-test inference on this model.

```
import numpy as np
from scipy.optimize import minimize
from scipy.stats import chi2

# Simulated data
np.random.seed(42)
N = 500
family_size = np.random.randint(1, 5, N)  # Instrument 1
non_labour_income = np.random.normal(20000, 5000, N)  # Instrument 2
education = np.random.randint(8, 20, N)  # Instrument 3
Z = np.column_stack((family_size, non_labour_income, education))  # Instrument matrix

# True parameters
theta_true = [0.05, 0.1]  # [theta1, theta2]

# Generate hours worked (endogenous variable)
hours = 40 + 2 * family_size - 0.001 * non_labour_income + 0.5 * np.random.normal(0, 1, N)

# Generate wages (dependent variable)
epsilon = np.random.normal(0, 100, N)
wage = np.exp(theta_true[0] * hours + theta_true[1] * education) + epsilon

# Define the moment conditions
def gmm_moments(theta, wage, hours, education, Z):
    theta1, theta2 = theta
    residual = wage - np.exp(theta1 * hours + theta2 * education)
    moments = Z.T @ residual / len(wage)
    return moments

# Define the GMM objective function
def gmm_objective(theta, wage, hours, education, Z, W):
    moments = gmm_moments(theta, wage, hours, education, Z)
    return moments.T @ W @ moments

# Initial parameter guesses
theta_init = [0.01, 0.01]

# Step 1: Use identity weighting matrix
W_identity = np.eye(Z.shape[1])
result_step1 = minimize(gmm_objective, theta_init, args=(wage, hours, education, Z, W_identity), method='BFGS')
theta_step1 = result_step1.x

# Step 2: Update the weighting matrix using estimated residuals
moments_step1 = gmm_moments(theta_step1, wage, hours, education, Z)
S_hat = np.cov(moments_step1)  # Variance of the moment conditions
W_optimal = np.linalg.inv(S_hat)  # Optimal weighting matrix

# Step 3: Re-estimate using optimal weighting matrix
result_step2 = minimize(gmm_objective, theta_step1, args=(wage, hours, education, Z, W_optimal), method='BFGS')
theta_step2 = result_step2.x

# Compute the J-statistic
moments_step2 = gmm_moments(theta_step2, wage, hours, education, Z)
J_stat = len(wage) * moments_step2.T @ W_optimal @ moments_step2  # J-statistic
df = Z.shape[1] - len(theta_step2)  # Degrees of freedom
p_value = 1 - chi2.cdf(J_stat, df)

# Print Results
print("Estimated Parameters (Two-Step GMM):")
print(f"Theta_1: {theta_step2[0]:.4f}, Theta_2: {theta_step2[1]:.4f}")
print("\nJ-Test for Overidentifying Restrictions:")
print(f"J-statistic: {J_stat:.4f}")
print(f"Degrees of Freedom: {df}")
print(f"P-value: {p_value:.4f}")
```