Econometrics 710
Final Exam
Spring, 2005
Sample Answers

1. The moment equations are

$$\begin{aligned} Eg_i(\mu) &= 0 \\ g_i(\mu) &= \begin{pmatrix} y_i - \mu \\ x_i \end{pmatrix} \end{aligned}$$

There are two moments, one parameter, so the model is overidentified. Let

$$\Omega = Eg_i g_i' = \begin{pmatrix} E(y_i-\mu)^2 & E(x_i(y_i-\mu)) \\ E(x_i(y_i-\mu)) & Ex_i^2 \end{pmatrix} = \begin{pmatrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}$$

where we make use of the knowledge that $Ex_i = 0$. Let

$$\overline{g}_n(\mu) = \frac{1}{n}\sum_{i=1}^n g_i(\mu) = \begin{pmatrix} \overline{y}_n - \mu \\ \overline{x}_n \end{pmatrix}$$

and

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_y^2 & \hat{\sigma}_{xy} \\ \hat{\sigma}_{xy} & \hat{\sigma}_x^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (y_i-\overline{y}_n)^2 & \frac{1}{n}\sum_{i=1}^n (y_i-\overline{y}_n)x_i \\ \frac{1}{n}\sum_{i=1}^n (y_i-\overline{y}_n)x_i & \frac{1}{n}\sum_{i=1}^n x_i^2 \end{pmatrix}$$

The efficient GMM estimator $\hat{\mu}$ for μ minimizes

$$\begin{aligned} J_n(\mu) &= n\overline{g}_n(\mu)'\Omega^{-1}\overline{g}_n(\mu) \\ &= \frac{n}{\hat{\sigma}_y^2\hat{\sigma}_x^2 - \hat{\sigma}_{xy}^2}\begin{pmatrix} \overline{y}_n - \mu & \overline{x}_n \end{pmatrix}\begin{pmatrix} \hat{\sigma}_x^2 & -\hat{\sigma}_{xy} \\ -\hat{\sigma}_{xy} & \hat{\sigma}_y^2 \end{pmatrix}\begin{pmatrix} \overline{y}_n - \mu \\ \overline{x}_n \end{pmatrix} \\ &= \frac{n}{\hat{\sigma}_y^2\hat{\sigma}_x^2 - \hat{\sigma}_{xy}^2}\left(\hat{\sigma}_x^2(\overline{y}_n-\mu)^2 - 2\hat{\sigma}_{xy}(\overline{y}_n-\mu)\overline{x}_n + \hat{\sigma}_y^2\overline{x}_n^2\right) \end{aligned}$$

The minimizer is

$$\hat{\mu} = \overline{y}_n - \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}\overline{x}_n. \tag{1}$$

Side Note: Interestingly, this is the same as the intercept from the OLS estimate of the equation

$$y_i = \hat{\mu} + \hat{\beta}x_i + e_i.$$

The important point is that the efficient GMM estimator (1) is not simply the sample mean $\overline{y}_n$. The latter is a GMM estimator, but it not efficient when we add the information that $Ex_i = 0$. (Unless $\sigma_{xy} = 0$, in which case the sample mean is efficient. However, this is not assumed in the question.)

2. Substituting $y_i = x_i'\beta + \varepsilon_i$, we obtain

$$\begin{aligned}
\tilde{\beta} - \beta &= \left(\sum_{i=1}^n \varepsilon_i^{-2} x_i x_i'\right)^{-1} \left(\sum_{i=1}^n \varepsilon_i^{-2} x_i \varepsilon_i\right) \\
&= \left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^{-2} x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^{-1} x_i\right)
\end{aligned}$$

By the WLLN,

$$\frac{1}{n}\sum_{i=1}^n \varepsilon_i^{-2} x_i x_i' \to_p E\left(\varepsilon_i^{-2} x_i x_i'\right) = Q,$$

$$\frac{1}{n}\sum_{i=1}^n \varepsilon_i^{-1} x_i \to_p E\left(\varepsilon_i^{-1} x_i\right) = \delta$$

In general, $\delta \neq 0$, so $\tilde{\beta} \to_p \beta + Q^{-1}\delta$. $\hat{\beta}$ is consistent for β iff $\delta = 0$. If ε_i is symmetric about zero, and $E\left|\varepsilon_i\right|^{-1} < \infty$ then $E\left(\varepsilon_i^{-1} \mid x_i\right) = 0$ and

$$\delta = E\left(\varepsilon_i^{-1} x_i\right) = E\left(x_i E\left(\varepsilon_i^{-1} \mid x_i\right)\right) = 0.$$

Furthermore, note that

$$E\left(\varepsilon_i^{-1} x_i - \delta\right)\left(\varepsilon_i^{-1} x_i - \delta\right)' = E\left(\varepsilon_i^{-2} x_i x_i'\right) - \delta\delta' = Q - \delta\delta'.$$

Then by the CLT

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\varepsilon_i^{-1} x_i - \delta\right) \to_d N\left(0, Q - \delta\delta'\right)$$

Thus

$$\sqrt{n}\left(\tilde{\beta} - (\beta + \delta)\right) \to_d N(0, V)$$

where

$$V = Q^{-1}\left(Q - \delta\delta'\right)Q^{-1} = Q^{-1} - Q^{-1}\delta\delta' Q^{-1}.$$

In the case where $\delta = 0$, this is

$$\sqrt{n}\left(\tilde{\beta} - \beta\right) \to_d N\left(0, Q^{-1}\right)$$

Infeasible GLS has the asymptotic distribution

$$\sqrt{n}\left(\tilde{\beta}_{GLS} - \beta\right) \to_d N\left(0, V_{GLS}\right)$$

$$V_{GLS} = \left(E\left(\sigma_i^{-2} x_i x_i'\right)\right)^{-1}$$

By Jensen's inequality

$$E\left(\varepsilon_i^{-2} \mid x_i\right) \geq \left(E\left(\varepsilon_i^2 \mid x_i\right)\right)^{-1} = \sigma_i^{-2}.$$

Therefore

$$Q = E\left(\varepsilon_i^{-2} x_i x_i'\right) = E\left(x_i x_i' E\left(\varepsilon_i^{-2} \mid x_i\right)\right) \geq E\left(x_i x_i' \sigma_i^{-2}\right)$$

and thus

$$V = Q^{-1} \leq E\left(x_i x_i' \sigma_i^{-2}\right)^{-1} = V_{GLS}$$

We can conclude that the infeasible estimator $\tilde{\beta}$ is more efficient than infeasible GLS (when $\delta = 0$). This seems impossible, as we know that GLS is asymptotically efficient. The trick is that there is no feasible version of $\tilde{\beta}$ which attains the same distribution, so the efficient gain is empirically irrelevant.

3. The model is just identified, so is estimated by OLS. Write the estimates as

$$y_i = x_{1i}' \hat{\beta}_1 + x_{2i}' \hat{\beta}_2 + \hat{e}_i$$

Let

$$\hat{V} = \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n x_i x_i' \hat{e}_i^2\right) \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1}.$$

The Wald statistic to test H_0 is

$$\begin{aligned} W_n &= n\left(\hat{\beta}_1 - \hat{\beta}_2\right)' \left(R'\hat{V}R\right)^{-1} \left(\hat{\beta}_1 - \hat{\beta}_2\right) \\ R &= \begin{pmatrix} I_k \\ -I_k \end{pmatrix} \end{aligned}$$

To do a nonparametric bootstrap test, we sample (y_i^*, x_i^*) jointly from the observations. On each bootstrap sample, we construct the OLS estimates

$$y_i^* = x_{1i}^{*\prime} \hat{\beta}_1^* + x_{2i}^{*\prime} \hat{\beta}_2^* + \hat{e}_i^*$$

covariance matrix

$$\hat{V}^* = \left(\frac{1}{n}\sum_{i=1}^n x_i^* x_i^{*\prime}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n x_i^* x_i^{*\prime} \hat{e}_i^{*2}\right) \left(\frac{1}{n}\sum_{i=1}^n x_i^* x_i^{*\prime}\right)^{-1}.$$

and Wald statistic

$$W_n^* = n\left(\left(\hat{\beta}_1^* - \hat{\beta}_2^*\right) - \left(\hat{\beta}_1 - \hat{\beta}_2\right)\right)' \left(R'\hat{V}^*R\right)^{-1} \left(\left(\hat{\beta}_1^* - \hat{\beta}_2^*\right) - \left(\hat{\beta}_1 - \hat{\beta}_2\right)\right)$$

It is very important that the statistic is centered at the sample values $\left(\hat{\beta}_1 - \hat{\beta}_2\right)$, rather than at the hypothesized value of 0. The estimated bootstrap p-value is the percentage of the simulated W_n^* that are largely than the sample value W_n. If there are B bootstrap replications, this is

$$p_n^* = \frac{1}{B}\sum_{b=1}^B 1\left(W_n^*(b) \geq W_n\right).$$

4. The GMM criterion is

$$\begin{aligned} J_n(\beta) &= n\overline{g}_n(\beta)'\hat{\Omega}^{-1}\overline{g}_n(\beta) \\ \overline{g}_n(\beta) &= \frac{1}{n}\left(X'Y - X'Z\beta\right) \\ \hat{\Omega} &= \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \hat{e}_i^2 \\ \hat{e}_i &= y_i - z_i'\hat{\beta} \end{aligned}$$

Let

$$\hat{\beta} = \left(Z'X\hat{\Omega}^{-1}X'Z\right)^{-1} Z'X\hat{\Omega}^{-1}X'Y$$

denote the unconstrained GMM estimator. The Lagrangian can be written as

$$J_n(\beta, \lambda) = \frac{1}{2n} J_n(\beta) - \lambda' R'\beta$$

where $\lambda \in R^q$ is a Lagrange multiplier. The factor $1/2n$ is unimportant but makes the calculations easier. The constrained estimator $(\tilde{\beta}, \tilde{\lambda})$ minimizes $J_n(\beta, \lambda)$. The first order conditions are

$$0 = \frac{\partial}{\partial \beta} J_n(\tilde{\beta}, \tilde{\lambda}) = -Z'X\hat{\Omega}^{-1}X'\left(Y - Z\tilde{\beta}\right) - R\tilde{\lambda} \tag{2}$$

$$0 = \frac{\partial}{\partial \lambda} J_n(\tilde{\beta}, \tilde{\lambda}) = R'\tilde{\beta}$$

Premultiply (2) by $\left(Z'X\hat{\Omega}^{-1}X'Z\right)^{-1}$ to obtain

$$\begin{aligned} \left(Z'X\hat{\Omega}^{-1}X'Z\right)^{-1} R\tilde{\lambda} &= -\left(Z'X\hat{\Omega}^{-1}X'Z\right)^{-1} Z'X\hat{\Omega}^{-1}X'\left(Y - Z\tilde{\beta}\right) \\ &= \tilde{\beta} - \hat{\beta}. \end{aligned} \tag{3}$$

Premultiplying by R', using $R'\tilde{\beta} = 0$, and solving,

$$\tilde{\lambda} = -\left(R'\left(Z'X\hat{\Omega}^{-1}X'Z\right)^{-1} R\right)^{-1} R'\hat{\beta}.$$

Substituting this into (3) we find

$$\tilde{\beta} = \hat{\beta} - \left(Z'X\hat{\Omega}^{-1}X'Z\right)^{-1} R\left(R'\left(Z'X\hat{\Omega}^{-1}X'Z\right)^{-1} R\right)^{-1} R'\hat{\beta}$$

5. Let

$$\begin{aligned} D_1 &= Z'P_1Z \\ D_2 &= Z'P_2Z \\ D_\lambda &= \lambda D_1 + (1-\lambda) D_2 \end{aligned}$$

Recall that

$$\begin{aligned} \hat{\beta}_1 &= D_1^{-1}Z'P_1Y \\ P_1 &= X_1 (X_1'X_1)^{-1} X_1' \\ \hat{\beta}_2 &= D_2^{-1}Z'P_2Y \\ P_2 &= X_2 (X_2'X_2)^{-1} X_2' \end{aligned}$$

and we calculate that

$$\begin{aligned} \tilde{\beta} &= (Z'XWX'Z)^{-1} (Z'XWX'Y) \\ &= \left(\begin{pmatrix} Z'X_1 & Z'X_2 \end{pmatrix} \begin{pmatrix} (X_1'X_1)^{-1}\lambda & 0 \\ 0 & (X_2'X_2)^{-1}(1-\lambda) \end{pmatrix} \begin{pmatrix} X_1'Z \\ X_2'Z \end{pmatrix} \right)^{-1} \\ &\quad \cdot \left(\begin{pmatrix} Z'X_1 & Z'X_2 \end{pmatrix} \begin{pmatrix} (X_1'X_1)^{-1}\lambda & 0 \\ 0 & (X_2'X_2)^{-1}(1-\lambda) \end{pmatrix} \begin{pmatrix} X_1'Y \\ X_2'Y \end{pmatrix} \right) \\ &= (\lambda Z'P_1Z + (1-\lambda) Z'P_2Z)^{-1} (\lambda Z'P_1Y + (1-\lambda) Z'P_2Y) \\ &= D_\lambda^{-1}\lambda Z'P_1Y + D_\lambda^{-1}(1-\lambda) Z'P_2Y \\ &= \lambda D_\lambda^{-1}D_1\hat{\beta}_1 + (1-\lambda) D_\lambda^{-1}D_2\hat{\beta}_2 \\ &= W_1\hat{\beta}_1 + W_2\hat{\beta}_2 \end{aligned}$$

where $W_1 = \lambda D_\lambda^{-1}D_1$ and $W_2 = (1-\lambda) D_\lambda^{-1}D_2$. $\tilde{\beta}$ is a weighted average since

$$W_1 + W_2 = D_\lambda^{-1}(\lambda D_1 + (1-\lambda) D_2) = I$$