Econometrics 710
Final Exam
Spring, 2007
Sample Answers

1.

(a) The sample moments are $\sum_{i=1}^n x_i\left(y_{1i}-x_i'\hat\beta_1\right)=0$ and $\sum_{i=1}^n x_i\left(y_{2i}-x_i'\hat\beta_2\right)=0$, which have solutions $\hat\beta_1=(X'X)^{-1}(X'Y_1)$ and $\hat\beta_2=(X'X)^{-1}(X'Y_2)$, which is equation-by-equation least squares.

(b) Writing the two regressions estimators in a vector,

$$\begin{aligned}\sqrt{n}\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_2-\beta_2\end{pmatrix} &= \begin{pmatrix}\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{\sqrt{n}}X'e_1\right)\\ \left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{\sqrt{n}}X'e_2\right)\end{pmatrix}\\ &= \begin{pmatrix}\left(\frac{1}{n}X'X\right)^{-1} & 0\\ 0 & \left(\frac{1}{n}X'X\right)^{-1}\end{pmatrix}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_i\right)\end{aligned}$$

where

$$u_i=\begin{pmatrix}x_ie_{1i}\\ x_ie_{2i}\end{pmatrix}.$$

From the WLLN and CLT,

$$\sqrt{n}\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_2-\beta_2\end{pmatrix}\to_d\begin{pmatrix}Q^{-1} & 0\\ 0 & Q^{-1}\end{pmatrix}N(0,\Omega)=N(0,V)$$

where $Q=Ex_ix_i'$,

$$\Omega=E\left(u_iu_i'\right)=\begin{pmatrix}E\left(x_ix_ie_{1i}^2\right) & E\left(x_ix_ie_{1i}e_{2i}\right)\\ E\left(x_ix_ie_{1i}e_{2i}\right) & E\left(x_ix_ie_{2i}^2\right)\end{pmatrix}=\begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{pmatrix}$$

and

$$V=\begin{pmatrix}Q^{-1} & 0\\ 0 & Q^{-1}\end{pmatrix}\Omega\begin{pmatrix}Q^{-1} & 0\\ 0 & Q^{-1}\end{pmatrix}=\begin{pmatrix}Q^{-1}\Omega_{11}Q^{-1} & Q^{-1}\Omega_{12}Q^{-1}\\ Q^{-1}\Omega_{21}Q^{-1} & Q^{-1}\Omega_{22}Q^{-1}\end{pmatrix}.$$

Note that the problem doesn't make any claim about the relationship between e_{1i} and e_{2i}, and therefore we need to allow them to be correlated, and should not assume that they are uncorrelated or independent.

(c) The hypothesis is linear, so the Wald statistic is appropriate to test H_0. The Wald statistic takes the form

$$W_n=n\left(\hat\beta_1-\hat\beta_2\right)'\hat V_\theta^{-1}\left(\hat\beta_1-\hat\beta_2\right)$$

where $\hat V_\theta$ is an estimate of V_θ, the asymptotic variance of $\sqrt{n}\left(\hat\beta_1-\hat\beta_2\right)$. There are two (equivalent) methods to find $\hat V_\theta$. The first is to note that

$$\begin{aligned}V_\theta &= \begin{pmatrix}I & -I\end{pmatrix}V\begin{pmatrix}I\\ -I\end{pmatrix}\\ &= Q^{-1}\begin{pmatrix}I & -I\end{pmatrix}\Omega\begin{pmatrix}I\\ -I\end{pmatrix}Q^{-1}\\ &= Q^{-1}\left(\Omega_{11}-\Omega_{21}-\Omega_{12}+\Omega_{22}\right)Q^{-1}\end{aligned}$$

and thus a natural estimate is

$$\hat{V}_\theta = \hat{Q}^{-1}\left(\hat{\Omega}_{11} - \hat{\Omega}_{21} - \hat{\Omega}_{12} + \hat{\Omega}_{22}\right)\hat{Q}^{-1}$$

where

$$\begin{aligned}
\hat{Q} &= \frac{1}{n}X'X \\
\hat{\Omega}_{11} &= \frac{1}{n}\sum_{i=1}^{n} x_i x_i \hat{e}_{1i}^2 \\
\hat{\Omega}_{11} &= \frac{1}{n}\sum_{i=1}^{n} x_i x_i \hat{e}_{1i}\hat{e}_{2i} \\
\hat{\Omega}_{22} &= \frac{1}{n}\sum_{i=1}^{n} x_i x_i \hat{e}_{2i}^2
\end{aligned}$$

and $\hat{e}_{1i} = y_{1i} - x_i'\hat{\beta}_1$ and $\hat{e}_{2i} = y_{2i} - x_i'\hat{\beta}_2$ are the OLS residuals. The second method to find $\hat{V}_\theta$ and W_n is to observe that

$$\begin{aligned}
\hat{\beta}_1 - \hat{\beta}_2 &= \left(X'X\right)^{-1}\left(X'Y_1\right) - \left(X'X\right)^{-1}\left(X'Y_2\right) \\
&= \left(X'X\right)^{-1}\left(X'\left(Y_1 - Y_2\right)\right)
\end{aligned}$$

Thus the hypothesis $\beta_1 = \beta_2$ is equivalent to the hypothesis of a zero coefficient vector in the regression of $y_1 - y_2$ on X. Algebraically, you'll obtain the same answer either way.

Since the Wald statistic W_n is asymptotically χ_k^2 under H_0, we reject H_0 in favor of H_1 when W_n exceeds the 5% critical value from the χ_k^2 distribution.

2. This is feasible since there are two parameters to estimate (β and γ) and two instruments (x_i and x_i^2). This is the just-identified case. Under the given assumtpion $E\left(e_i \mid x_i\right) = 0$, we know that $E\left(x_i e_i\right) = 0$ and $E\left(x_i^2 e_i\right) = 0$, so these are valid instrumental variables. For identification, we need that the included endogenous variable (z_i) be correlated with the excluded exogenous variabe (x_i^2) after controlling for the included exogenous variable (x_i). That is, in the reduced form regression

$$z_i = x_i\alpha_1 + x_i^2\alpha_2 + u_i \tag{1}$$

it must be that $\alpha_2 \neq 0$. Thus the conditional mean of z_i given x_i cannot be linear, it must be a non-linear relationship. Note: This is not the same thing $E\left(z_i x_i^2\right) \neq 0$, but it is close.

In summary, the proposed GMM estimator is valid, if coefficient α_2 in the reduced form equation (1) is non-zero.

3. The efficient GMM estimator is

$$\begin{aligned}
\hat{\beta} &= \left(X'\begin{pmatrix} X & Q \end{pmatrix}\hat{\Omega}^{-1}\begin{pmatrix} X' \\ Q' \end{pmatrix}X\right)^{-1}\left(X'\begin{pmatrix} X & Q \end{pmatrix}\hat{\Omega}^{-1}\begin{pmatrix} X' \\ Q' \end{pmatrix}Y\right) \\
&= \left(\begin{pmatrix} X'X & X'Q \end{pmatrix}\hat{\Omega}^{-1}\begin{pmatrix} X'X \\ Q'X \end{pmatrix}\right)^{-1}\left(\begin{pmatrix} X'X & X'Q \end{pmatrix}\hat{\Omega}^{-1}\begin{pmatrix} X'Y \\ Q'Y \end{pmatrix}\right)
\end{aligned}$$

where

$$\begin{aligned}
\hat{\Omega} &= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_i \\ q_i \end{pmatrix}\begin{pmatrix} x_i' & q_i' \end{pmatrix}\hat{e}_i^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_i x_i' & x_i q_i' \\ q_i x_i' & q_i q_i' \end{pmatrix}\hat{e}_i^2
\end{aligned}$$

and $\hat{e}_i = y_i - x_i'\tilde{\beta}$ with $\tilde{\beta}$ a preliminary consistent estimator. One possibility is $\tilde{\beta} = (X'X)^{-1}(X'Y)$, the LS estimator.

4. The method of moments estimator for σ^2 is

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2 \\ \hat{e}_i &= y_i - x_i'\hat{\beta} \\ \hat{\beta} &= (X'X)^{-1}(X'Y).\end{aligned}$$

The bootstrap percentile interval is constructed as follows. Let $\{y_i^*, x_i^* : i = 1, ..., n\}$ denote a random sample of size n drawn from the observations. The estimator is then applied to the bootstrap data:

$$\begin{aligned}\hat{\sigma}^{*2} &= \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^{*2} \\ \hat{e}_i^* &= y_i^* - x_i^{*\prime}\hat{\beta}^* \\ \hat{\beta}^* &= (X^{*\prime}X^*)^{-1}(X^{*\prime}Y^*).\end{aligned}$$

This is repeated B times. (B is a large number, typically $B = 999$ or higher.) We now have a psuedo-sample of B draws from the distribution of $\hat{\sigma}^{*2}$. Let $\hat{\sigma}_b^{*2}$ denote the bth replication. The standard bootstrap $(1-\alpha)\%$ percentile interval is $[\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}]$, where $\hat{q}_a$ is the $a'th$ empirical quantile of the psuedo-sample $\{\hat{\sigma}_1^{*2}, ..., \hat{\sigma}_B^{*2}\}$. Numerically, this is found by sorting the $\hat{\sigma}_b^{*2}$ values. If $B = 999$ and $\alpha = 0.05$ these are the 25'th and 975'th sorted values.

5.

(a) By the WLLN and substituting $y_i = x_i^*\beta + e_i = v_i^{-1}x_i\beta + e_i$, and using the fact that e_i is independent of x_i and mean zero,

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}x_iy_i\right) \\ &\xrightarrow{p} \left(Ex_i^2\right)^{-1}E(x_iy_i) \\ &= \frac{E\left(x_i\left(v_i^{-1}x_i\beta + e_i\right)\right)}{E\left(x_i^2\right)} \\ &= \frac{E\left(v_i^{-1}x_i^2\right)}{E\left(x_i^2\right)}\beta\end{aligned}$$

This is the plim for $\hat{\beta}$, expressed in terms of moments of x_i and v_i. This is what the question asked. My intention had been to express the plim in terms of moments of x_i^* and v_i, which is more insightful. Since $x_i^2 = x_i^{*2}v_i^2$ and x_i^* and v_i are independent, we can write the plim as

$$= \frac{E\left(x_i^{*2}v_i\beta\right)}{E\left(x_i^{*2}v_i^2\right)} = \left(\frac{E(v_i)}{E\left(v_i^2\right)}\right)\beta$$

This also can be written as $\beta\mu_v/\left(\mu_v^2 + \sigma_v^2\right)$, where μ_v and σ_v^2 are the mean and variance of v_i.

(b) $\hat{\beta}$ is consistent only if $\left(\frac{Ev_i}{Ev_i^2}\right)\beta = \beta$, which requires either that $\beta = 0$, or $Ev_i = Ev_i^2$. The latter is equivalent to the condition $\mu_v = \mu_v^2 + \sigma_v^2$. This cannot happen if $\mu_v \geq 1$, but is possible if $\mu_v < 1$. The measurement error v_i takes a proportionate form. It would be reasonable to describe the measurement error v_i as unbiased if $Ev_i = 1$ or $E\ln v_i = 0$. (Note: these are not the same!) Both situations are incompatible with $Ev_i = Ev_i^2$, as they are incompatible with $\mu_v < 1$.