Econ 203c, Problem Set 5, Spring 2003

Question 1

1) Definitions

a) A sequence of random variables, X_n, converges to μ almost surely if for all $\epsilon > 0$,

$pr(\lim_{n\to\infty} |X_n - \mu| > \epsilon) = 0$

b) A sequence of random variables, X_n, converges to μ in probability if for all $\epsilon > 0$,

$\lim_{n\to\infty} pr(|X_n - \mu| > \epsilon) = 0$

2) Example of a sequence of random variables that converges in probability but does not converge almost surely.

Casella and Berger (1990), p.215

Let $s \in [0,1]$ and $X(s) = s$.

The sequence of random variables $X_1, X_2, \ldots$ is defined according to this pattern

$X_1(s) = s + I\{0 \leq s \leq 1\}$

$X_2(s) = s + I\{0 \leq s \leq 1/2\}$

$X_3(s) = s + I\{1/2 \leq s \leq 1\}$

$X_4(s) = s + I\{0 \leq s \leq 1/3\}$

$X_5(s) = s + I\{1/3 \leq s \leq 2/3\}$

$X_6(s) = s + I\{2/3 \leq s \leq 1\}$

and so on...

a) Convergence in probability:

It's fairly easy to see that as $\lim_{n\to\infty} pr(|X_n - X| > \epsilon) = 0$.

b) But this sequence does not converge almost surely. There is at least some s at which this sequence does not converge to X. There is no pointwise convergence.

$pr(\lim_{n\to\infty} |X_n - X| > \epsilon) > 0$.

Question 2

Show: $\widehat{\theta}_n \xrightarrow{p} \theta_0$

$$\widehat{\theta}_n = \arg\min m_n(\theta)' V_n^{-1} m_n(\theta),$$

where

$V_n \xrightarrow{p} V$, $m_n(\theta) = 1/n \sum \varphi(y,x,\theta)$, and $E_0[\varphi(y,x,\theta_0)] = 0$.

Proof:

The easiest way to prove this is to directly use the MLE proof in the lecture by redefining the GMM objective function.

Let $Q_n(\theta) = -m_n(\theta)' V_n^{-1} m_n(\theta)$.

Then, $\widehat{\theta}_n = \arg\max Q_n(\theta)$.

Now the GMM estimator maximizes this $Q_n(\theta)$ function. This corresponds to the MLE, which maximises the log-likelihood. (This is the joke Moshe told you about how mathematicians solve problems by moving the tea kettle into the kitchen and starting over...or something).

There are three immediate implications of the model and the definition of the GMM estimator:

First, the definition of $\widehat{\theta}_n$ implies,

$$Q_n(\widehat{\theta}_n) \geq Q_n(\theta_0). \qquad (1)$$

Second, because $Q(\bullet)$ is the negative of a quadratic form, for any θ,

$$Q(\theta) \leq 0. \qquad (2)$$

Third, since θ_0 is the population parameter, for any θ,

$$Q(\theta) \leq Q(\theta_0). \qquad (3)$$

Using the assumptions above, law of large numbers, and Slutsky,

$$Q_n(\theta_0) \xrightarrow{p} E_0[\varphi(y,x,\theta_0)]' V^{-1} E_0[\varphi(y,x,\theta_0)] = Q(\theta_0).$$

By the definition of convergence in probability, there exists a sequence of non-negative random variables $\{Z_n\}$, with $Z_n \overset{p}{\to} 0$, and

$$|Q_n(\theta_0) - Q(\theta_0)| \leq Z_n. \quad (4)$$

Subtracting $Q(\widehat{\theta}_n)$ from both sides of (1), we can rewrite this as

$$Q(\theta_0) - Q(\widehat{\theta}_n) \leq Q_n(\theta_0) - Q(\widehat{\theta}_n) + Z_n.$$

$$\leq \sup_{\theta\in\Theta}|Q_n(\theta) - Q(\theta)| + Z_n, \quad (5)$$

The right hand side converges in probability to zero

$$\sup_{\theta\in\Theta}|Q_n(\theta) - Q(\theta)| + Z_n \overset{p}{\to} 0.$$

This implies that the left hand side also converges to zero since $Q(\theta_0) - Q(\widehat{\theta}_n) \geq 0$, by (2) and (3).

$$Q(\theta_0) - Q(\widehat{\theta}_n) \overset{p}{\to} 0. \quad (6)$$

There exists some $\eta > 0$, such that we can rewrite (3) as

$$Q(\theta) \leq Q(\theta_0) - \eta.$$

If we consider only $\theta \in \Theta$ s.t. $|\theta - \theta_0| \geq \epsilon$, for some $\epsilon > 0$, then this equality becomes strict.

$$Q(\theta) < Q(\theta_0) - \eta. \quad (7)$$

Therefore,

$$pr(|\widehat{\theta}_n - \theta_0| \geq \epsilon) \leq pr(Q(\widehat{\theta}_n) < Q(\theta_0) - \eta),$$

where the probability on the right hand side converges to zero by (6) and (7).

Question 3

Model: $y_i = x_i'\beta_0 + \epsilon_i$

$E[y_i|x_i] = x_i'\beta_0$

Note: since x is kx1, these moment conditions should be writen as follows–different than in the problem set:

Define: $\varphi_1(y,x,\beta) = x(y - x'\beta)$ and $\varphi_2(y,x,\beta) = z(y - x'\beta)$

where $z = (x_1,\ldots,x_K,x_2,\ldots,x_K^2)'$

1) $\beta_0 = \dfrac{\partial E[y_i|x_i]}{\partial x_i'}$

2) $E[\varphi_1(y,x,\beta_0)]$
$= E[x(y - x'\beta_0)]$
$= E_x\{E[x(y - x'\beta_0)|x]\}$

$E[x(y - x'\beta_0)|x] = x(E[y|x] - x'\beta_0) = 0$ since $E[y|x] = x'\beta_0$.
Thus, $E[\varphi_1(y,x,\beta_0)] = 0$.

Same argument for the second set of moments since z is a function of the xs.
$E[\varphi_2(y,x,\beta_0)]$
$= E[z(y - x'\beta_0)]$
$= E_x\{E[z(y - x'\beta_0)|x]\}$
$E[z(y - x'\beta_0)|x] = z(E[y|x] - x'\beta_0) = 0$ since $E[y|x] = x'\beta_0$.

Thus, $E[\varphi_2(y,x,\beta_0)] = 0$.

3) Optimal GMM estimator
$\widehat{\beta}_n^1 = \arg\min m_n^1(\beta)' V_{n,1}^{-1} m_n^1(\beta)$

where $m_n^1(\beta) = 1/n \sum x_i(y_i - x_i'\beta)$, $V_{n,1}(\beta_*^1) = 1/n \sum\{[x_i(y_i - x_i'\beta_*^1)][x_i(y_i - x_i'\beta_*^1)]'\}$, and β_*^1 is a consistent estimate for β_0 (e.g. obtained from an initial estimate using $\varphi_1(y,x,\beta_0)$ and $V = I$).

$\widehat{\beta}_n^2 = \arg\min m_n^2(\beta)' V_{n,2}^{-1} m_n^2(\beta)$

where $m_n^2(\beta) = 1/n \sum z_i(y_i - x_i'\beta)$, $V_{n,2}(\beta_*^2) = 1/n \sum\{[z_i(y_i - x_i'\beta_*^2)][z_i(y_i - x_i'\beta_*^2)]'\}$, and β_*^2 is a consistent estimate for β_0 (e.g. obtained from an initial estimate using $\varphi_2(y,x,\beta_0)$ and $V = I$).

4) Asymptotic covariance matrices
$\sqrt{n}\,(\widehat{\beta}_n^1 - \beta_0) \xrightarrow{d} N(0, \Lambda_1(\beta_0))$, where
$\Lambda_1(\beta_0) = (A_1(\beta_0) V_1^{-1} A_1(\beta_0)')^{-1}$

and $A_1(\beta_0) = \dfrac{\partial E[\varphi_1(y,x,\beta_0)]}{\partial \beta} = x'x$ and $V_1 = E[\varphi_1(y,x,\beta_0)\varphi_1(y,x,\beta_0)']$

$\sqrt{n}\,(\widehat{\beta}_n^2 - \beta_0) \xrightarrow{d} N(0, \Lambda_2(\beta_0))$, where
$\Lambda_2(\beta_0) = (A_2(\beta_0) V_2^{-1} A_2(\beta_0)')^{-1}$
and $A_2(\beta_0) = \dfrac{\partial E[\varphi_2(y,x,\beta_0)]}{\partial \beta} = z'x$ and $V_2 = E[\varphi_2(y,x,\beta_0)\varphi_2(y,x,\beta_0)']$

5) Consistent estimators for the asymptotic covariance matrices.

First set of moments:
$\widehat{\Lambda}_1(\widehat{\beta}_n^1) = (\widehat{A}_1(\widehat{\beta}_n^1) V_{n,1}^{-1}(\widehat{\beta}_n^1) \widehat{A}_1(\widehat{\beta}_n^1)')^{-1}$
where $\widehat{A}_1(\widehat{\beta}_n^1) = \dfrac{\partial m_n^1(\widehat{\beta}_n^1)}{\partial \beta}$, $m_n^1(\cdot)$ defined above, and $V_{n,1}(\cdot)$ defined above.

Consistency: $\widehat{\Lambda}_1(\widehat{\beta}_n^1) \xrightarrow{p} \Lambda_1(\beta_0)$
Proof:

By weak law of large numbers, $\widehat{A}_1(\beta_0) \xrightarrow{p} \dfrac{\partial E[\varphi_1(y,x,\beta_0)]}{\partial \beta}$.

If we assume that $\widehat{A}_1(\beta) \xrightarrow{p} \dfrac{\partial E[\varphi_1(y,x,\beta)]}{\partial \beta}$ uniformly over β,
then $\widehat{A}_1(\widehat{\beta}_n^1) \xrightarrow{p} A_1(\beta_0)$

By weak law of large numbers, $V_{n,1}(\beta_0) \xrightarrow{p} E[\varphi_1(y,x,\beta_0)\varphi_1(y,x,\beta_0)']$.
If we assume that $V_{n,1}(\beta) \xrightarrow{p} E[\varphi_1(y,x,\beta)\varphi_1(y,x,\beta)']$ uniformly over β, then
$V_{n,1}(\beta_*^1) \xrightarrow{p} V_1$

Thus, by Slutsky,
$$(\widehat{A}_1(\widehat{\beta}_n^1) V_{n,1}^{-1}(\widehat{\beta}_n^1) \widehat{A}_1(\widehat{\beta}_n^1)')^{-1} \xrightarrow{p} (A_1(\beta_0) V_1^{-1} A_1(\beta_0)')^{-1}$$

Second set of moments:
$\widehat{\Lambda}_2(\widehat{\beta}_n^2) = (\widehat{A}_2(\widehat{\beta}_n^2) V_{n,2}^{-1}(\widehat{\beta}_n^2) \widehat{A}_2(\widehat{\beta}_n^2)')^{-1}$
where $\widehat{A}_2(\widehat{\beta}_n^2) = \dfrac{\partial m_n^2(\widehat{\beta}_n^2)}{\partial \beta}$, $m_n^2(\cdot)$ defined above, and $V_{n,2}(\cdot)$ defined above.

Consistency: $\widehat{\Lambda}_2(\widehat{\beta}_n^2) \overset{p}{\to} \Lambda_2(\beta_0)$.

Same proof as above.

6) Asymptotic distribution using $\varphi_2(y,x,\beta) = (y - x'\beta)z$ and $V = I$
$\sqrt{n}\,(\widehat{\beta}_n^2 - \beta_0) \overset{d}{\to} N(0, \Lambda_2(\beta_0))$,

where $\Lambda_2(\beta_0) = (A_2(\beta_0)A_2(\beta_0)')^{-1}A_2(\beta_0)W_2(\beta_0)A_2(\beta_0)'(A_2(\beta_0)A_2(\beta_0)')^{-1}$, $A_2(\beta_0)$ defined above, and $W_2(\beta_0) = E_0[\varphi_2(y,x,\beta_0)\varphi_2(y,x,\beta_0)']$

Question 4:

It's a little unclear the exact order of this question,
but there are four cases for estimating β_0:

case i) Use $\varphi_1(y,x,\beta)$ and $V_{n,1}^{-1} = I$
In this case, $\widehat{\beta}_n^1 = \arg\min m_n^1(\beta)'m_n^1(\beta)$.

FOC: $\dfrac{\partial Q_n^1(\beta)}{\partial\beta} = 0$
$\Rightarrow 2\dfrac{\partial m_n^1(\beta)}{\partial\beta'}m_n^1(\beta) = 0$
$\Rightarrow (1/n\sum x_i x_i')(1/n\sum x_i(y_i - x_i'\beta)) = 0$
$\Rightarrow (x'x)[x'y - x'x\beta] = 0$
$\Rightarrow x'y - x'x\beta = 0$
$\Rightarrow \widehat{\beta}_n^1 = (x'x)^{-1}x'y$
Estimator for asymptotic covariance:
$\widehat{\Lambda_1}(\widehat{\beta}_n^1) = 1/n(\widehat{A_1}(\widehat{\beta}_n^1)\widehat{A_1}(\widehat{\beta}_n^1)')^{-1}\widehat{A_1}(\widehat{\beta}_n^1)\widehat{W}_1(\widehat{\beta}_n^1)A_1(\widehat{\beta}_n^1)'(A_1(\widehat{\beta}_n^1)A_1(\widehat{\beta}_n^1)')^{-1}$,
where $\widehat{A}_1(\widehat{\beta}_n^1) = \dfrac{\partial m_n^1(\widehat{\beta}_n^1)}{\partial\beta}$ and $\widehat{W}_1(\widehat{\beta}_n^1) = 1/n\sum\{[x_i(y_i - x_i'\widehat{\beta}_n^1)][x_i(y_i - x_i'\widehat{\beta}_n^1)]'\}$.

Collecting terms: $\widehat{\Lambda_1}(\widehat{\beta}_n^1) = n\{[(x'x)(x'x)']^{-1}(x'x)\widehat{W}_1(\widehat{\beta}_n^1)(x'x)'[(x'x)(x'x)']^{-1}$,
since $\widehat{A}_1(\widehat{\beta}_n^1) = 1/nx'x$

case ii) Use $\varphi_1(y,x,\beta)$ and $V_{n,1}(\beta_*^1) = 1/n\sum\{[x_i(y_i - x_i'\beta_*^1)][x_i(y_i - x_i'\beta_*^1)]'\}$ (optimal weight matrix)

Since this is the just identified case, the weight matrix does not affect the numerical estimator: $\widehat{\beta}_n^1 = (x'x)^{-1}x'y$.

Since it does not affect the numerical estimator, it also does not affect the estimated covariance of the estimator.

(In the Matlab code, I calculate it anyway and show this.)

case iii) Use $\varphi_2(y,x,\beta)$ and $V_{n,2}^{-1} = I$

Because this is over-identified, there is no exact solution to $m_n^2(\beta) = 0$ as above.
FOC: $\frac{\partial Q_n^2(\beta)}{\partial \beta} = 0$

$$\Rightarrow 2\frac{\partial m_n^2(\beta)}{\partial \beta'} m_n^2(\beta) = 0$$
$$\Rightarrow (1/n\sum x_i z_i')(1/n\sum z_i(y_i - x_i'\beta) = 0$$
$$\Rightarrow (x'z)[(z'y) - (z'x)\beta] = 0$$
$$\Rightarrow (x'z)(z'y) - (x'z)(z'x)\beta = 0$$
$$\Rightarrow \widehat{\beta}_n^2 = [(x'z)(z'x)]^{-1}(x'z)(z'y)$$

The estimated covariance matrix is
$\widehat{\Lambda}_2(\widehat{\beta}_n^2) = n[(x'z)(z'x)]^{-1}(x'z)\widehat{W}_2(\widehat{\beta}_n^2)(z'x)[(x'z)(z'x)]^{-1}$,
where $\widehat{W}_2(\widehat{\beta}_n^2) = 1/n\sum\{[z_i(y_i - x_i'\widehat{\beta}_n^2)][z_i(y_i - x_i'\widehat{\beta}_n^2)]'\}$.

case iv) Use $\varphi_2(y,x,\beta)$ and $V_{n,2}(\beta_*^2) = 1/n\sum\{[z_i(y_i - x_i'\beta_*^2)][z_i(y_i - x_i'\beta_*^2)]'\}$ (optimal weight matrix)

Because this is over-identified the weight matrix will affect the point estimate of β_0.
FOC: $\frac{\partial Q_n^2(\beta)}{\partial \beta} = 0$

$$\Rightarrow 2\frac{\partial m_n^2(\beta)}{\partial \beta'} V_{n,2}^{-1}(\beta_*^2) m_n^2(\beta) = 0$$

Hence, $\widehat{\beta}_n^2 = [(x'z)V_{n,2}^{-1}(\beta_*^2)(z'x)]^{-1}(x'z)V_{n,2}^{-1}(\beta_*^2)(z'y)$.

The estimated asymptotic covariance matrix is

$$\begin{aligned}\widehat{\Lambda}_2(\widehat{\beta}_n^2) &= 1/n(\widehat{A}_2(\widehat{\beta}_n^2)V_{n,2}^{-1}(\widehat{\beta}_n^2)\widehat{A}_2(\widehat{\beta}_n^2)')^{-1} \\ &= n[(x'z)V_{n,2}^{-1}(\widehat{\beta}_n^2)(z'x)]^{-1},\end{aligned}$$

where $V_{n,2}^{-1}(\widehat{\beta}_n^2) = 1/n\sum\{[z_i(y_i - x_i'\widehat{\beta}_n^2)][z_i(y_i - x_i'\widehat{\beta}_n^2)]'\}$

Results (Matlab code attached):
(numbers refer to cases defined above)

beta_gmm1 =
1.0169
1.0004
-1.0194
0.4928
-0.4896
se1 =
0.0978
0.0236
0.0229
0.0220
0.0253
beta_gmm2 =
1.0169
1.0004
-1.0194
0.4928
-0.4896
se2 =
0.0978
0.0236
0.0229
0.0220
0.0253
beta_gmm3 =
1.1686
0.9878

-1.0459
0.4873
-0.5135
se3 =
0.1577
0.0309
0.0298
0.0281
0.0349
beta_gmm4 =
1.0177
0.9961
-1.0168
0.4903
-0.4875
se4 =
0.0978
0.0233
0.0227
0.0219
0.0246

3) and 4) Discussion

The standard errors and point estimates for the just identified cases (case i and case ii) are exactly the same.

For the over-identified case, there is an advantage (smaller standard errors) to using the optimal weight matrix.

5) Hypothesis testing

$$H_0 : \beta_2^2 + \beta_4^2 - \beta_3^2 - \beta_5^2 = 0$$
$$H_1 : \beta_2^2 + \beta_4^2 - \beta_3^2 - \beta_5^2 \neq 0$$

Wald statistic:

Using $\varphi_1(y,x,\beta)$, the Wald statistic is

$$W_1 = \frac{(\widehat{\beta_2^2} + \widehat{\beta}_4^2 - \widehat{\beta}_3^2 - \widehat{\beta}_5^2)^2}{\Psi_1(\widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4, \widehat{\beta}_5)},$$

where $\Psi_1(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4, \widehat{\beta}_5) = \Delta_1 \widehat{\Lambda}_1(\widehat{\beta}_n^1)\Delta_1'$, all the point estimates are from the first set of moments, and $\Delta_1 = [0, 2\widehat{\beta}_2, -2\widehat{\beta}_3, 2\widehat{\beta}_4, -2\widehat{\beta}_5]$

Results: $W_1 =$0.2293

Using $\varphi_2(y,x,\beta)$, the Wald statistic is

$$W_2 = \frac{(\widehat{\beta_2^2} + \widehat{\beta}_4^2 - \widehat{\beta}_3^2 - \widehat{\beta}_5^2)^2}{\Psi_2(\widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4, \widehat{\beta}_5)},$$

where $\Psi_2(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4, \widehat{\beta}_5) = \Delta_2 \widehat{\Lambda}_2(\widehat{\beta}_n^2)\Delta_2'$, all the point estimates are from the first set of moments, and $\Delta_2 = [0, 2\widehat{\beta}_2, -2\widehat{\beta}_3, 2\widehat{\beta}_4, -2\widehat{\beta}_5]$

Results: $W_2 =$0.2859

$W_1 \sim \chi(1)$ and $W_2 \sim \chi(1)$

Conclusion: At a significance level of 95 percent, the critical value is 3.84. Since both Wald statistics are very close and less than this critical value, I would fail to reject the null hypothesis using either set of moments.