

## Solutions to Assessed Problem Set 2

1 (i). We can write

$$G_T(\hat{\theta}_T) - G_0 = \{G_T(\hat{\theta}_T) - G(\hat{\theta}_T)\} + \{G(\hat{\theta}_T) - G(\theta_0)\} \quad (1)$$

It follows from (1) via the Triangle Inequality - see Practice Problem set #4 Question 4(b) - that

$$\|G_T(\hat{\theta}_T) - G_0\| \leq \|G_T(\hat{\theta}_T) - G(\hat{\theta}_T)\| + \|G(\hat{\theta}_T) - G(\theta_0)\| \quad (2)$$

1(ii) As stated in the question, under Assumptions 3.1-3.4 and 3.7-3.10, we can invoke Theorem 3.1 to deduce that  $\hat{\theta}_T \xrightarrow{p} \theta_0$ , or in other words that

$$\lim_{T \rightarrow \infty} P(\|\hat{\theta}_T - \theta_0\| < \delta) = 1 \quad (3)$$

for all  $\delta > 0$ , and hence that

$$\lim_{T \rightarrow \infty} P(\hat{\theta}_T \in N_\epsilon) = 1 \quad (4)$$

where  $N_\epsilon$  denotes an  $\epsilon$  neighbourhood of  $\theta_0$  for any  $\epsilon > 0$ . Therefore, we can restrict attention to the  $\epsilon$  - neighbourhood defined in Assumption 3.13. By definition of the sup, it follows that for  $\hat{\theta}_T \in N_\epsilon$ ,

$$\|G_T(\hat{\theta}_T) - G(\hat{\theta}_T)\| \leq \sup_{\theta \in N_\epsilon} \|G_T(\theta) - G(\theta)\| \quad (5)$$

Assumption 3.13 states that

$$\sup_{\theta \in N_\epsilon} \|G_T(\theta) - G(\theta)\| \xrightarrow{p} 0 \quad (6)$$

and it follows from (5)-(6) that

$$\|G_T(\hat{\theta}_T) - G(\hat{\theta}_T)\| = o_p(1) \quad (7)$$

1(iii). Again, it suffices to restrict attention to  $\hat{\theta}_T \in N_\epsilon$  (for the reasons given above) and we choose  $\epsilon$  such that  $G(\theta)$  is continuous on  $N_\epsilon$  (see Assumption 3.12). Assumption 3.12 implies that for every  $\eta$ , there exists  $0 < \delta \leq \epsilon$  such that

$$|G_{i,j}(\hat{\theta}_T) - G_{i,j}(\theta_0)| < \eta, \quad \text{for } \|\theta - \theta_0\| < \delta \quad (8)$$

where  $G_{i,j}(\theta)$  is the  $i - j$  element of  $G(\theta)$ . Since (3) holds for all  $\delta > 0$ , it follows from (8) that

$$\lim_{T \rightarrow \infty} P(|G_{i,j}(\hat{\theta}_T) - G_{i,j}(\theta_0)| < \eta) = 1 \quad (9)$$

for all  $\eta$  such that  $0 < \eta \leq \epsilon$ , or, in other words that,  $G_{i,j}(\hat{\theta}_T) - G_{i,j}(\theta_0) = o_p(1)$ . Since (9) holds for all  $i, j$ , it follows - via Practice Problem Set #4 Question 5 - that

$$\|G(\hat{\theta}_T) - G(\theta_0)\| = o_p(1) \quad (10)$$

1(iv) From the definition of the norm and parts (i)-(iii), it follows that  $0 < \|G_T(\hat{\theta}_T) - G_0\| \leq \xi_T$  where  $\xi_T = o_p(1)$ . It therefore follows that  $\|G_T(\hat{\theta}_T) - G_0\| = o_p(1)$  and hence (from Practice Problem Set Question 4(a)) that  $G_T(\hat{\theta}_T) - G_0 = o_p(1)$  which is the desired result.

2(i) Since  $v_t$  is iid, it follows that the  $i - j^{\text{th}}$  element of  $S$  is

$$S_{i,j} = E[f_i f_j] \quad (11)$$

where  $f_j = f_j(v_t, \theta_0)$  is the  $j^{\text{th}}$  element of  $f(v_t, \theta_0)$ . To evaluate the expectation of these terms, we need to calculate the probabilities of certain events. By definition of the model,  $P(D_t(i) = 1) = h(i; \theta_0) = h_i$ . We also need  $P(D_t(i) = a, D_t(j) = b)$  for  $a, b \in \{0, 1\}$ . These probabilities are as follows:

- $P(D_t(i) = 1, D_t(j) = 1)$ : since  $D_t(i) = 1$  and  $D_t(j) = 1$  are mutually exclusive events for  $i \neq j$ , it follows that

$$P(D_t(i) = 1, D_t(j) = 1) = 0 \quad (12)$$

- $P(D_t(i) = 1, D_t(j) = 0)$ : since  $D_t(i) = 1$  can only happen with  $D_t(j) = 0$ , it follows that  $P(D_t(i) = 1, D_t(j) = 0) = P(D_t(i) = 1)$  and so

$$P(D_t(i) = 1, D_t(j) = 0) = h_i \quad (13)$$

- $P(D_t(i) = 0, D_t(j) = 0)$ : if neither  $D_t(i)$  or  $D_t(j)$  is one then one of the other indicator variables must be one and so

$$P(D_t(i) = 0, D_t(j) = 0) = 1 - h_i - h_j \quad (14)$$

Using these probabilities we have

$$\begin{aligned} \text{Var}[f_i] &= (1 - h_i)^2 h_i + (0 - h_i)^2 (1 - h_i) \\ &= h_i (1 - h_i) \end{aligned} \quad (15)$$

$$\begin{aligned} \text{Cov}[f_i, f_j] &= (1 - h_i)(-h_j)h_i + (-h_i)(1 - h_j)h_j + (-h_i)(-h_j)(1 - h_i - h_j) \\ &= -h_i h_j \end{aligned} \quad (16)$$

The desired result then follows from (11), (15) and (16).

2(ii) Recall that a matrix is singular if at least one column can be written as a linear combination of the others. Furthermore one column must be a linear combination of the others if the sum of the columns is zero. Letting  $S_{\cdot,j}$  be the  $j^{\text{th}}$  column of  $S$ , we

$$\sum_{j=1}^k S_{\cdot,j} = \begin{bmatrix} \sum_{j=1}^k S_{1,j} \\ \sum_{j=1}^k S_{2,j} \\ \cdot \\ \cdot \\ \sum_{j=1}^k S_{k,j} \end{bmatrix} \quad (17)$$

Now, for any  $i$ , we have

$$\sum_{j=1}^k S_{i,j} = h_i(1 - \sum_{j=1}^k h_j) = 0 \quad (18)$$

because  $\sum_{i=1}^k h_i = 1$  by construction.

2(iii) We prove this is the appropriate inverse by considering the product of the two matrices. Let  $a_i$  be the  $i^{\text{th}}$  row of  $S_1$  and  $b_j$  be the  $j^{\text{th}}$  column of  $S_*$ . From the definition of the inverse, it follows that to establish that  $S_* = S^{-1}$  we must show that  $a'_i b_i = 1$  and  $a'_i b_j = 0$  for  $i \neq j$ . Multiplying out we have:

$$\begin{aligned} a'_i b_i &= h_i(1 - h_i)\left(\frac{1}{h_i} + \frac{1}{h_k}\right) + \sum_{s \neq i}^{k-1} \frac{-h_i h_s}{h_k} \\ &= (1 - h_i) + \frac{h_i[(1 - \sum_{s=1}^{k-1} h_s)]}{h_k} \\ &= 1 - h_i + \frac{h_i[h_k]}{h_k} \\ &= 1 \end{aligned} \quad (19)$$

and

$$\begin{aligned} a'_i b_j &= h_i(1 - h_i)\frac{1}{h_k} - h_i h_j\left(\frac{1}{h_j} + \frac{1}{h_k}\right) + \sum_{s \neq i,j}^{k-1} \frac{-h_i h_s}{h_k} \\ &= \frac{h_i}{h_k} \left(1 - h_i - h_j - \sum_{s \neq i,j}^{k-1} h_s\right) - h_i \\ &= \frac{h_i(h_k)}{h_k} - h_i \\ &= 0 \end{aligned} \quad (20)$$

2(iv) The GMM estimation based on  $E[f_1(v_t, \theta_0)] = 0$  with weighting matrix  $W_T = \hat{S}_*$  has minimum

$$Q_T(\theta) = T^{-1} \sum_{t=1}^T f_1(v_t, \theta)' \hat{S}_* T^{-1} \sum_{t=1}^T f_1(v_t, \theta) \quad (21)$$

Note that

$$T^{-1} \sum_{t=1}^T f_1(v_t, \theta) = \begin{bmatrix} \hat{p}_1 - h_1(\theta) \\ \hat{p}_2 - h_2(\theta) \\ \vdots \\ \hat{p}_{k-1} - h_{k-1}(\theta) \end{bmatrix} \quad (22)$$

where  $\hat{p}_i = T^{-1} \sum_{t=1}^T D_t(i)$  and  $h_i(\theta) = h(i; \theta)$ . Substituting (22) into (21) and using the definition of  $\hat{S}_*$  given in the question, it follows that

$$\begin{aligned}
Q_T(\theta) &= \sum_{i=1}^{k-1} \left( \frac{[\hat{p}_i - h_i(\theta)]^2}{\hat{p}_i} \right) + \frac{1}{\hat{p}_k} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} [\hat{p}_i - h_i(\theta)][\hat{p}_j - h_j(\theta)] \\
&= \sum_{i=1}^{k-1} \left( \frac{[\hat{p}_i - h_i(\theta)]^2}{\hat{p}_i} \right) + \frac{1}{\hat{p}_k} \left\{ \sum_{i=1}^{k-1} [\hat{p}_i - h_i(\theta)] \right\}^2 \\
&= \sum_{i=1}^k \left( \frac{[\hat{p}_i - h_i(\theta)]^2}{\hat{p}_i} \right)
\end{aligned} \tag{23}$$

where the last equality uses the fact that  $\sum_{i=1}^k [\hat{p}_i - h_i(\theta)] = 0$  and so  $\sum_{i=1}^{k-1} [\hat{p}_i - h_i(\theta)] = -[\hat{p}_k - h_k(\theta)]$ . It follows from (23) that  $TQ_T(\theta) = GF_T(\theta)$ .

- 3(v) From part (iv), it follows that  $\tilde{\theta}_T = \hat{\theta}_T$  where  $\hat{\theta}_T = \operatorname{argmin}_{\theta \in \Theta} Q_T(\theta)$ . To complete the proof, it must be shown that  $\hat{S}_* \xrightarrow{p} S_1^{-1}$  so that  $Q_T(\theta)$  is a GMM minimand with a weighting matrix that converges in probability to the inverse of the long run variance of the  $T^{-1/2} \sum_{t=1}^T f_1(v_t, \theta_0)$ , and hence that  $TQ_T(\hat{\theta}_T)$  is asymptotically equivalent to the overidentifying restrictions test based on  $E[f_1(v_t, \theta_0)] = 0$ . Since  $v_t$  is iid, we can use the Weak Law of Large Numbers to deduce

$$\hat{p}_i = T^{-1} \sum_{t=1}^T D_t(i) \xrightarrow{p} E[D_t(i)] = h_i \tag{24}$$

It then follows from Slutsky's Theorem that  $\hat{S}_* \xrightarrow{p} S_1^{-1}$  because the elements of the inverse are continuous functions of the elements of the original matrix.