

Uitwerkingen Tentamen Econometrie 3, FEE-UvA  
21 oktober 2002, 9:30-11:30 uur, gebouw A, zaal C

1. Beschouw het regressie model met slechts één verklarende variabele  $y_i = x_i\beta + \varepsilon_i$ , waarbij  $E\{\varepsilon_i | X\} = 0$ ,  $E\{\varepsilon_i^2 | X\} = \sigma_i^2 > 0$  en  $E\{\varepsilon_i\varepsilon_j | X\} = 0$  voor  $i \neq j = 1, \dots, N$ , met  $X = (x_1, \dots, x_N)'$ .

(a) {5} Voor de schatter  $b$  geldt  $b = \beta + \sum_{i=1}^N x_i\varepsilon_i / \sum_{j=1}^N x_j^2$  zodat  $E\{b | X\} = \beta + \left(\sum_{i=1}^N x_i^2\right)^{-1} \sum_{i=1}^N x_i E\{\varepsilon_i | X\} = \beta$ . Dus  $b$  is zuiver, want  $E\{b\} = EE\{b | X\} = E\beta = \beta$ .

(b) {5}  $V\{b | X\} = E\{(b - \beta)^2 | X\} = \left(\sum_{i=1}^N x_i^2\right)^{-2} \sum_{i=1}^N \sum_{j=1}^N x_i x_j E\{\varepsilon_i \varepsilon_j | X\}$ , dus  $V\{b | X\} = \left(\sum_{i=1}^N x_i^2\right)^{-2} \sum_{i=1}^N \sigma_i^2 x_i^2$ .

(c) {10} De hypothese  $H_0 : \sigma_i^2 = \sigma^2, \forall i$  kan getoetst worden met de White toets. Men doet dan de hulpregressie  $e_i^2 = \alpha + x_i\delta + x_i^2\gamma + error$ , waarbij  $e_i = y_i - x_i b$ . De toetsgrootte is dan  $NR^2$  (waarbij  $R^2$  de centered  $R^2$  van de hulpregressie is) en deze is asymptotisch  $\chi^2(2)$  verdeeld, dus bij een (asymptotisch) significantie niveau van 5% verwerpt men  $H_0$  als  $NR^2 > \chi_{0.95}^2(2)$ , maar er zijn meer mogelijkheden.

(d) {5} De heteroskedasticiteit consistente covariantie schatter voor  $V\{b | X\}$  luidt voor dit model  $\hat{V}\{b | X\} = \sum_{i=1}^N e_i^2 x_i^2 / \left(\sum_{i=1}^N x_i^2\right)^2$ , waarbij  $e_i = y_i - x_i b$ . De gebruikelijke OLS covariantieschatter is  $s^2 / \sum_{i=1}^N x_i^2$ , waarbij  $s^2 = \frac{1}{N-1} \sum_{i=1}^N e_i^2$  omdat  $K = 1$ , dus voor het model met alleen een constante ( $x_i \equiv 1, \forall i$ ) is dat  $\frac{1}{N(N-1)} \sum_{i=1}^N e_i^2$ , terwijl dan  $\hat{V}\{b | X\} = \sum_{i=1}^N e_i^2 x_i^2 / \left(\sum_{i=1}^N x_i^2\right)^2 = \frac{1}{N^2} \sum_{i=1}^N e_i^2$ , dus op de vrijheidsgraden correctie na zijn ze dan gelijk.

(e) {15} Met Monte Carlo onderzoek kan men op de volgende manier verifiëren of bij een bepaalde gegeven  $x_i\beta$  en  $\sigma_i^2$  voor  $i = 1, \dots, N$  geldt  $E\{\hat{V}\{b | X\}\} \neq V\{b | X\}$ . We voeren  $R$  replicaties  $r = 1, \dots, R$  uit, waarbij we genereren  $y_i^{(r)} = x_i\beta + \varepsilon_i^{(r)}$ , met  $\varepsilon_i^{(r)} \sim N(0, \sigma_i^2)$ . In de opgave staat niet uit welke verdeling de storingen komen, dus daar moeten we zelf iets voor kiezen. Dat maakt onze resultaten wel specifiek voor die keuze. Door iedere replicatie OLS uit te voeren vinden we  $b^{(r)}$  en  $e_i^{(r)} = y_i^{(r)} - x_i b^{(r)}$  en kunnen we uitrekenen  $v^{(r)} = \sum_{i=1}^N \left(e_i^{(r)}\right)^2 x_i^2 / \left(\sum_{i=1}^N x_i^2\right)^2$ . De Monte Carlo schatter van  $E\{\hat{V}\{b | X\}\}$  is dan  $\hat{v} = \frac{1}{R} \sum_{r=1}^R v^{(r)}$  en de standaardfout van die schatter is  $se(\hat{v}) = \frac{1}{\sqrt{R}} \sqrt{\frac{1}{R-1} \sum_{r=1}^R (v^{(r)} - \hat{v})^2}$ , dus een (omtrekt 95%) betrouwbaarheidsinterval voor  $E\{\hat{V}\{b | X\}\}$  is  $\hat{v} \pm 2se(\hat{v})$ . De populatie waarde  $V\{b | X\}$  kunnen we exact uitrekenen volgens (b) of uit de Monte Carlo schatten op grond van de  $b^{(r)}$  met  $\frac{1}{R-1} \sum_{r=1}^R (b^{(r)} - \bar{b})^2$ , waarbij  $\bar{b} = \frac{1}{R} \sum_{r=1}^R b^{(r)}$ . Hoe groter  $R$  hoe beter de Monte Carlo schatters. Als (de Monte Carlo schatter van)  $V\{b | X\}$  buiten het interval  $\hat{v} \pm 2se(\hat{v})$  ligt dan blijkt dat  $E\{\hat{V}\{b | X\}\} \neq V\{b | X\}$ . Kennelijk is  $N$  dan niet groot genoeg om al zichtbaar te maken dat  $\text{plim}_{N \rightarrow \infty} \hat{V}\{b | X\} = V\{b | X\}$ .

2. Een onderzoeker heeft lineaire regressie gedaan van een variabele  $y_i$  op een  $K \times 1$  vector van regressoren  $x_i$ . Hij twijfelt over de gekozen functionele vorm en besluit een RESET toets uit te voeren.

(a) {5} De RESET hulpregressie luidt  $y_i = x_i'\beta + \hat{y}_i^2\gamma_2 + \dots + \hat{y}_i^Q\gamma_Q + error$ , waarbij  $\hat{y}_i = x_i'b$ . Er moet dus nog een keuze omtrent  $Q$  gemaakt worden, waarbij  $2 \leq Q \ll N - K$ .

- (b) {10} De toets op  $\gamma_2 = \dots = \gamma_Q = 0$  wordt in de hulpregressie uitgevoerd met een  $F$  toets. Laat de residuele kwadraten som van het oorspronkelijke model  $RSS$  zijn, en van de hulpregressie  $HRSS$ , dan luidt de toetsgrootheid  $\frac{N-K-Q+1}{Q-1} \frac{RSS-HRSS}{HRSS}$  en de relevante kritieke waarde is  $F_{Q-1, N-K-Q+1}(1-\alpha)$ . Maar, men kan minder subtiel te werk gaan en het Wald principe rechtstreeks toepassen. De toetsgrootheid is dan  $N \frac{RSS-HRSS}{HRSS}$  en de relevante kritieke waarde  $\chi_{Q-1}^2(1-\alpha)$ .
- (c) {10} Als de RESET toets een  $p$ -waarde van 0.09 oplevert, dan betekent dat het volgende. Laat de  $F$ -toetsgrootheid de waarde  $w$  aan hebben genomen. Dan geldt  $P\{F_{Q-1, N-K-Q+1} > w\} = 0.09$ . Dus een trekking uit de  $F$  verdeling met de betreffende vrijheidsgraden is met een kans van 0.09 groter dan de gevonden waarde  $w$ . Levert de Wald vorm bij een waarde  $w$  een  $p$ -waarde van 0.09 op dan geldt  $P\{\chi_{Q-1}^2 > w\} = 0.09$ .
- (d) {5} We staan voor de beslissing de huidige functionele vorm wel of niet te verwerpen, en daarbij de kans beperkt te houden dat we ten onrechte verwerpen, maar ook weer niet zo beperkt dat we ook bij een verkeerde functionele vorm maar met geringe kans de huidige specificatie verwerpen. Ten onrechte verwerpen lijkt hier minder erg dan tenonrechte niet verwerpen, dus een significantie niveau van 5% is aan de lage kant. Maar, hanteer je dat, dan geeft een  $p$ -waarde van 0.09 geen aanleiding tot verwerpen. Kies je voor een significantie niveau van 10% dan verwerp je bij een  $p$ -waarde van 0.09 wel.
- (e) {5} De RESET toets is geen exacte toets, omdat niet geldt  $E\{\varepsilon_i | \hat{y}_i^j\} = 0$ . Alleen asymptotisch zal het nominale significantie niveau gelijk zijn aan de werkelijke kans op type I fouten. Bij een eindige steekproef weten we dus niet precies tot welke waarde we de kans op type I fouten beperken.
3. Bij het binaire keuze model specificeert men  $P\{y_i = 1\} = F(x_i'\beta)$  en  $P\{y_i = 0\} = 1 - F(x_i'\beta)$ , waarbij  $\beta$  en  $x_i$  allebei niet-stochastische  $K \times 1$  vectoren zijn.
- (a) {10} De achterliggende overwegingen hierbij zijn dat er een latente (niet waargenomen) variabele  $y_i^*$  is waarvoor  $y_i^* = x_i'\beta + \varepsilon_i$ , met  $\varepsilon_i$  een symmetisch om 0 verdeelde storing, terwijl  $y_i = 1$  als  $y_i^* > 0$  en  $y_i = 0$  als  $y_i^* \leq 0$ . Er geldt dan  $P\{y_i = 1\} = P\{y_i^* > 0\} = P\{x_i'\beta + \varepsilon_i > 0\} = P\{\varepsilon_i > -x_i'\beta\} = P\{\varepsilon_i < x_i'\beta\} = F(x_i'\beta)$  en  $P\{y_i = 0\} = 1 - F(x_i'\beta)$ . Daarbij moet de functie  $F$  de kenmerken van een verdelingsfunctie hebben, dus  $f(v) = \frac{\partial F(v)}{\partial v} \geq 0$  en  $F(-\infty) = 0$  en  $F(+\infty) = 1$ .
- (b) {5}  $L(\beta) = \prod_{i=1}^N P\{y_i = 1\}^{y_i} P\{y_i = 0\}^{1-y_i}$  zodat  $\log L(\beta) = \sum_{i=1}^N \{y_i \log F(x_i'\beta) + (1 - y_i) \log[1 - F(x_i'\beta)]\}$ .
- (c) {5} Voor de score vector  $s(\beta) = \frac{\partial \log L(\beta)}{\partial \beta}$  geldt
- $$\begin{aligned} s(\beta) &= \sum_{i=1}^N \left\{ y_i \frac{\partial}{\partial \beta} \log F(x_i'\beta) + (1 - y_i) \frac{\partial}{\partial \beta} \log[1 - F(x_i'\beta)] \right\} \\ &= \sum_{i=1}^N \left\{ y_i \frac{f(x_i'\beta)}{F(x_i'\beta)} + (1 - y_i) \frac{-f(x_i'\beta)}{1 - F(x_i'\beta)} \right\} x_i \\ &= \sum_{i=1}^N \frac{y_i f(x_i'\beta) - y_i f(x_i'\beta) F(x_i'\beta) - f(x_i'\beta) F(x_i'\beta) + y_i f(x_i'\beta) F(x_i'\beta)}{F(x_i'\beta)[1 - F(x_i'\beta)]} x_i \\ &= \sum_{i=1}^N \frac{y_i - F(x_i'\beta)}{F(x_i'\beta)[1 - F(x_i'\beta)]} f(x_i'\beta) x_i \end{aligned}$$
- (d) {5} De score vector heeft verwachting nul, want  $E\{y_i\} = 1 \times P\{y_i = 1\} + 0 \times P\{y_i = 0\} = F(x_i'\beta)$ .