KTH Teknikvetenskap
Harald Lang

Tentamensskrivning, 12/03-2009, kl 14.00–19.00.
SF2951 Ekonometri
Short answers

1. The "normal equations" says that $\sum_1^n x_i\hat{e}_i = 0$. Note that $\hat{y}_i = x_i'\hat{\beta} = \hat{\beta}' x_i$. We get

$$\begin{aligned}\sum_1^n y_i^2 &= \sum_1^n (\hat{\beta}' x_i + \hat{e}_i)^2 = \sum_1^n (\hat{\beta}' x_i)^2 + 2\sum_1^n \hat{\beta}' x_i \hat{e}_i + \sum_1^n \hat{e}_i^2 \\ &= \sum_1^n \hat{y}_i^2 + 0 + \sum_1^n \hat{e}_i^2\end{aligned}$$

Q.E.D.

2. Let y_i = dummy for "not repaid", $x_{1,i}$ = dummy for "record of non-payment during the last two years", $x_{2,i}$ = dummy for "record of non-payment during the period two to five years back in time" and $x_{3,i}$ = "interest rate of the loan". Here i refers to borrower number i. We model the probability $p(x_1, x_2, x_3)$ as a "Logit":

$$p(x_1, x_2, x_3) = \frac{\exp(\beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3)}{1 + \exp(\beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3)}$$

and either estimate this equation by ML, or the equation

$$y = \frac{\exp(\beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3)}{1 + \exp(\beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3)} + \mathit{error}$$

by GNLLS (known covariance matrix).

3a. There are more than one possible test, but a Wald test seems most natural.

b. If the null is true, then the Wald test says that

$$\frac{1.48^2}{0.45} + \frac{1.36^2}{0.38} \approx 9.73$$

is an observation of a $\chi^2(2)$-variable. Since this exceeds 9.21, we reject the null at the 1% level.

4a. The interpretation of $x'\beta$ is $x'\beta = \text{median}(y \mid x)$.

b. Possible reasons include: we *want* to estimate the median rather than the expected value of y, the estimation is more robust to outliers, invariance to monotonic transformations.

5a. Solve for the reduced form. We get

$$y_2 = \frac{1}{1 - \alpha_2\beta_2}([\beta_0 + \beta_2\alpha_0] + \beta_2\alpha_1 x_1 + \beta_1 x_2 + [e_2 + \beta_2 e_1])$$

so we see that y_2 is correlated with e_1 (unless $\beta_2 = 0$.)

b. Both equations can be estimated with IV (2SLS) with the exogeneous variables x_1 and x_2 as instruments.

6. We re-estimate β N times ($N = 2'000$ or some such) using resampling of with replacement of n observations. Denote the new estimates $\hat{\beta}_i^*,\ i = 1\ldots,n$. Now compute the values $g(x_0;\hat{\beta}_i^*),\ i = 1\ldots,n$, and choose a and b such that 0.5% of the values $g(x_0;\hat{\beta}_i^*)$ falls above b, and 0.5% of the values $g(x_0;\hat{\beta}_i^*)$ falls below a. Now an approximate 1% conficence interval for $g(x_0;\beta)$ is

$$2g(x_0;\hat{\beta}) - b \leq g(x_0;\beta) \leq 2g(x_0;\hat{\beta}) - a.$$

Here $\hat{\beta}$ is the original estimate. Similarly for any other level of conficence.