AREC422 Notes

Youpei Yan

April 11, 2019

Review Session

1. Functional forms and related issues

· Scaling: t and F tests, R^2 and adjusted R^2 are not affected.

· Z-score's interpretation: x_j increases by 1 sd, $\hat{y}$ changes by $\hat{\beta}_j sd$.

· Logarithmic: $\widehat{log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ...$, then $\Delta\widehat{log(y)} = \hat{\beta}_1 \Delta x_1$:
Approximation: $100\hat{\beta}_1\Delta x_1\%$; Exact Change: $100 \cdot [exp(\hat{\beta}_1 \Delta x_1) - 1]\%$.
e.g.: Approximate method works well when the change is relatively small: If $\widehat{log(y)} = 0.05$, then we have 5% and 5.13% as the approximate change and the exact change in y. If $\widehat{log(y)} = 0.25$, then we have 25% and 28.4% as the approximate change and the exact change in y.

· Quadratics: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$, then the turning point $x^* = |\hat{\beta}_1/(2\hat{\beta}_2)|$.
In reality, include the quadratic term in the model if the corresponding coefficient is significant.

· Partial Effect (PE) of x: $\Delta y/\Delta x$, or dy/dx.
e.g. $log(price) = \beta_0 + \beta_1 size + \beta_2 size^2 + \beta_3 log(NO_2) + \beta_4 log(NO_2)^2 + u$, then:
the PE of size $dlog(price)/dsize = \beta_1 + 2\beta_2 size$,
the exact impact of price: $\underline{100}[exp(\beta_1 + 2\beta_2 size) - 1]\%$, is a function of size.
the APE of size: $\beta_1 + 2\beta_2 \overline{size}$, we use $mean(size)$ to find the value.
the PE of $log(NO_2)$ $dlog(price)/dlog(NO_2) = \beta_3 + 2\beta_4 log(NO_2) =$ elasticity of price with respect to NO_2, which is also a function of NO_2.

· Interaction Terms: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$, then
the PE of $x_2 = \beta_2 + \beta_3 x_1$, which means that an additional level of x_2 yields a higher increase in y for higher x_1 if $\beta_2 + \beta_3 x_1 > 0$ and $\beta_3 > 0$.
the APE of $x_2 = \hat{\beta}_2 + \hat{\beta}_3 \bar{x}_1$

· Adjusted R^2: correction for R^2 by imposing a penalty for adding more control variables. It is useful to choose between non-nested models. (Recall that nested models' comparison is done by using an F-test.)

· CI for predictions: with values for $(x_1, x_2, ...x_k)$ are given.
2 types of CIs: 1. predicted value of y and 2. any future value of y, with the given characteristics.
Method: run the regression $y = \theta_0 + \beta_1(x_1 - c_1) + \beta_2(x_2 - c_2) + ... + \beta_k(x_k - c_k) + u$, and get the estimate and the standard error of $\hat{\theta}$.
We need the residual standard error $\hat{\sigma}$ for the second CI.
2 types of CIs are equal to: 1. $[\hat{\theta} \pm 1.96 \cdot se(\hat{\theta})]$ and 2. $[\hat{\theta} \pm 1.96 \cdot \sqrt{se^2(\hat{\theta}) + \hat{\sigma}^2}]$

II. Qualitative Inference

A. Dummy variable on the RHS
zero-one variable: qualitative information rather than quantitative information.

1. $y = \beta_0 + \beta_1 x_1 + \delta_0 d + u$
We have 3 ways of interpreting δ_0:
in words: the difference in y between one group ($d = 1$) and the other ($d = 0$).
math: $\delta_0 = E(y|d = 1, Xs) - E(y|d = 0, Xs)$, conditional on Xs, the only difference in y is caused by the dummy variable.
graph: two parallel lines with the same slope as β_1, but different intercepts: β_0 and $(\beta_0 + \delta_0)$.
e.g. If $y = log(price)$, when $d = 1$ is comparing to $d = 0$, price will increase by $100[exp(\hat{\delta}_0) - 1]\%$.

2. Ordinal Info with categorical variable
Why can't we include the categorical variable directly?
How many dummies should we create? (category-1) to avoid the perfect collinearity. And the omitted group is called the base group.
$y = \beta_0 + \delta_1 cate\_1 + \delta_2 cate\_2 + \delta_3 cate\_3 + \beta_k X_k + u$, the base group is $cate\_0$.
δ_2: when other factors are fixed, the difference in y between Category 2 and Category 0.
If y is pollution level, $cate_i$s are choices of abatement, $\delta_2 < 0$ means a better abatement choice comparing to Category 0 because the level of pollution is lower.
If we change the model and omit $cate\_2$, then
$y = \beta_0 + \delta_1' cate\_1 + \delta_2' cate\_0 + \delta_3 cate\_3 + \beta_k X_k + u$, the base group becomes $cate\_2$.
And we have $\delta_2' = -\delta_2$, they estimate the same difference, but we switch the position of the 2 groups.

* Chow Test among groups:
With m groups, UR model:
$y_1 = \beta_{01} + \beta_{11} x_1 + \beta_{21} x_2 + ... \beta_{k1} x_k + u$, with n_1 obs.
$y_2 = \beta_{02} + \beta_{12} x_1 + \beta_{22} x_2 + ... \beta_{k2} x_k + u$, with n_2 obs.
...
$y_m = \beta_{0m} + \beta_{1m} x_1 + \beta_{2m} x_2 + ... \beta_{km} x_k + u$, with n_m obs.
R model:
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$, with $n = n_1 + n_2 + ... + n_m$ obs.
\# Restrictions $= q = (m-1)(k+1)$ with slopes and intercept restrictions.
\# Restrictions $= q = (m-1)k$ with just slopes restrictions.
$df_{ur} = (n_1 + n_2 + ... + n_k) - m(k+1)$

$$F_{q,[n-m(k+1)]} = \frac{[SSR_r - (SSR_1 + SSR_2 ... + SSR_m)]/q}{(SSR_1 + SSR_2 ... + SSR_m)/[n - m(k+1)]}$$

B. Dummy Variable on the LHS
Binary response model: $P(y = 1|\mathbf{X}) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$
$\hat{\beta}_j$: increasing in x_j, change in the predicted probability of "success", when other factors are fixed.
LPM's shortcomings: outliers, heteroscedasticity.
We usually use logit and probit.

Generally, any models can fall into OVB issues. Remember the formula $E(\tilde{\beta}_i) = E(\hat{\beta}_i + \hat{\beta}_j \tilde{\delta})$, where $\tilde{\delta}$ is the correlation between x_i and x_j.