Econometrics: Methods and Applications Cousera

linear regression

numeric y => OLS

Specification y = bX+e

Assumptions

OLS assumption
A1 Linear model: y = X β + ε.
A2 Fixed regressors: X non-random.
A3 Random error terms with mean zero: E (ei) = 0.
A4 Homoskedastic error terms: E (ε2i) = σ2 for all i = 1, . . . , n.
A5 Uncorrelated error terms: E(εiεj) = 0 for all i ≠ j.
A6 Parameters β and σ2 are fixed and unknown.

Interpretation

Total effect = Partial effect + Indirect effect
dy / dx_j
=∂y / ∂x_j +sum(∂y / ∂x_h*∂x_h /∂x_h)
if x_h ! = x_j

- Partial effect: ∂y/∂x_j = βj if x_h remains fixed for all h != j
- Indirect effect: sum(∂y / ∂x_h*∂x_h /∂x_h) if x_h ! = x_j

Transformation

common transformations

- difference — Trend pattern
- log — Exponential growth
- interaction terms — Xi*Xj
- higher order terms
- multiple levels dummy variables

Model specification
Select b

Iteractive Methods
- backward elimination
- foward elimination

Selection Critiera
- T-test, F-test
- Ramsey RESET test: test on non linear effect
- in-sample information criteria
  - aic: aic = log(s^2) + 2k/n; s: standard error of OLS; k: # of variables
  - bic: bic = log(s^2) + klog(n)/n; s: standard error of OLS; k: # of variables

Model Evaluation

Residuals diagnostic

Common methods
- residual scatter plot
- Jarque-Bera Test: Error Normality test

Common residual problems
http://docs.statwing.com/interpreting-residual-plots-to-improve-your-regression/#transform
- Y-axis Unbalanced
- Large Y-axis Datapoints
- Nonlinear pattern
- X-axis Unbalanced
- Outliers
- Heteroscedasticity

out-of-sample
- RMSE
- MAE
- parameter stability
  - chow-break — parameter consistency between in-sample and out-sample
  - chow forward

Multicollinearity

correlation != causility
Endogeneity

diagnose on 2SLS

is X endogenous?
examine the following reasons
- Omitted variables
- Strategic behavior
- Measurement errors

2SLS procedure
example:
x1 potentially endogenous variables
X2 exogenous variables
Z = (Z*, X2) instruments
First stage regression:
X1 = Z*r1 + X2 r2 + e

example:
x1 potentially endogenous variables
X2 exogenous variables
Z = (Z*, X2) instruments
2nd stage regression:
y = b1 X1 + X2 b2 + e
X1 is the regression result from step 1

test instrument Z
- enough instruments — more than # of endogenous variables
- Z correlated with X??? first stage 2SLS, t and f test
  example:
  x1 potentially endogenous variables
  X2 exogenous variables
  Z = (Z*, X2) instruments
  First stage regression:
  X1 = Z*r1 + X2 r2 + e
  Test H0: r1 = 0 in first -stage
- z not correlated with e??? sagan test
  1. USE Z to obtain 2SLS estimator b_2SLS
  2. e_2SLS = y - Xb_2SLS
  3. Regress e_2SLS on Z
  4. nR^2 ~ X^2(m-k) under H_0 (valid instruments m instruments in Z and k explanatory variables in X

Hausman test
Ingredients:
Explanatory variables: X = (X1, X2)
Potentially endogenous: X1(k1 variables)
Exogenous variables: X2(k2 variables)
Instrument:Z
NUll hypothies (H0): X1 is exogenous
1. regress y on X -> e = y - Xb
2. regress X1 on Z -> calculate residuals V
3. regress e on X+ V
4. nR2~ X^2(k1)
- reject — 2SLS
- not reject — OLS

dummy y -> Logit model MLE

Specification y = exp(bX)/(1+exp(bX) + e

interpretation
- odd ratio: Pr(y = 1)/ Pr(y =0) = exp(β1+β2*xi)
- Marginal effect: d Pr(y = 1)/ d x_i = Pr(y=1)Pr(y=0)*b_i

Logistic Diagnostic test
- pseudo R2
  - McFadden R2
  - Nagelkerke R2
- Confusion matrix — Precision and recall
- signle restriction — T-test
- multiple restrictions — χ2 test

Time Series

Intuition

Past values often have predictive power for future

Stationary for each individual time series?
y, X1, X2, X3, ...

ADF test with trend
Δyt =α+pyt−1 +γ1Δyt−1 +...+γLΔyt−L +εt
set lag p by `AIC', `BIC', `t-stat'

ADF test without trend
Δyt =α+pyt−1 +γ1Δyt−1 +...+γLΔyt−L +εt
set lag p by `AIC', `BIC', `t-stat'

go to

Non-stationarity not rejected
- take Δyt as new y

Non-stationarity rejected

VAR

Specification
- yt = α + β1yt−1 + β2yt−2 + . . . + βpyt−p + εt
- y_t = F(PY_{t-1}, PX(t-1))
  PY_{t-1} = (y_{t-1}, y_{t-2},....,y_{1})

Time series diagnostic test
- choice of lag length — aic, bic
- stability test — chow break test; chow forward test
- normal residuals — jarque-bera test
- uncorrelated residuals
  - abs(ACF) > 1/ sqrt(n)
  - breusch-Godfrey test
    example
    Model yt = α + βyt−1 + γxt−1 + εt
    → Step 1: OLS residuals et = yt − a − byt−1 − cxt −1.
    → Step 2: OLS in et = α + βyt−1 + γxt−1 + δ1et−1 + δ2et−2 + ut
    → Step 3: BG = nR2 ≈ χ2(2) if et white noise.
    → Conclusion: Model not correctly specified if BG > 6.0.
    → Should then adjust model, e.g. more lags of yt and xt
- out-of-sample test
  - RMSE
  - MAE
  - SUM

select mutually linked x with y

Granger causality test
-T-test
-F-test coefficients jointly

cointegrated between x and y?

Engle-Granger test
1. OLS in yt = α + βxt + εt → b and residuals e
2. Cointegrated if ADF test on e_t
rejects non-stationarity
Δet = α + pe_{t-1} + γ1Δe_{t-1} + . . . + γLΔet-L + ω

cointegration

Error Correction Model (ECM)
- specification
  - Δyt = α + β1(yt−1 − bxt−1) + β2Δyt−1 + β3Δxt−1 + εt
  - Δyt = α + β1ecm_{t-1} + β2Δyt−1 + β3Δxt−1 + εt
- **Time series diagnostic test**

reject cointegration

ADL
- **Time series diagnostic test**
- Specification: yt = α + β1yt−1 + β2yt−2 + . . . + βpyt−p + + rlxt_1 + r2 xt-2 + rrx t-r εt

Statistical Distributions

- normal distribution
- multivarate normal distribution
- student-t
- F-distribution
- χ2 (chi-square) distribution