SAS Mathi

- Panel data ⇒ unobserved heterogeneity → creates bias
- Cross sections: firm & household
- Time series: TS: weekly, hourly
- bias & consistency
- Observed heterogeneity

② adjusted Req
⑥ R-square

① interaction
② own price elast.
③ cross price elast
④ seasonality effect
⑤ Competition effect

- Complex world with multiple segments
- hierarchical model : ~~un~~observed : Interaction check

$$Sales = \alpha + \beta p + (\alpha_1 Inc + \alpha_2 Age + \alpha_3 Age^2)$$

$$(\beta_1 Inc + \beta_2 Age + \beta_3 Age^2)$$

- degree of freedom $\alpha + 100$ dummies for α_i when $N=100$

$$S_{it} = \alpha_i + \beta . price_{it}$$

- α_t → check versus effect of trend
  more than two ~~bias~~ obs from each person (remove bias)

✓ Cross sec Ver > time series Ver

- Random effect: $S_{it} = \alpha_i + \beta P_{it} + \varepsilon_{it}$
  $\alpha_i = \alpha + \xi_i \rightarrow \alpha, \sigma_{\xi_i}$
  Random distrib
  Assump: no Correl
  { ξ_i & ε_i not correlated
  { ξ_i & P_{it} "
  Could be: $\alpha_i = \bar{\alpha} + \xi_i + v_t$: $\sigma_{\xi_i}, \sigma_{v_t}$, $\bar{\alpha}$

Proc TSCSReg → ① CS ② TS
Panel
mixed

Housman test: fixed effect if Rejected

identification problem

① hetrogeneity
② Endogeneity
- ~~pre~~ predetermined
- Instrument variable Correlation with dependent through Endogeneous Var

Proc Syslin

Hetrodesticity — (logit, probit, discrim anal)
non linear estimations

- Proc model - Constant Ver assump across obs

① - white test (GLM)
  $y^2, y, e^y, \sqrt{y}$, $\ln y$, $1/y$ ε/x_i
② Gold Fild Quant (GQ test)
WLS

✱ influence → talks about outlier
DEFITS $> 2\sqrt{\frac{p}{n}}$ → #param, #obs
DF-Betas [] $\frac{2}{\sqrt{n}}$ Cmp
Outlier

✱ multiCollinearity → There is no ... in data
$y = a + b x_1 + b x_2$: x_1, x_2 Collin
biased estimate
⇒ Var β↑ ⇒ t-val ↓ (effect while assume not)
drop one variable
① check correl
② VIF >10
③ Coeff ver > 0.9
Ⓘ principal component analysis
Ⓘ Ridge Regression
eign value
VIF Collin

① unbiasedness $E(\beta) = \beta$
② Efficiency: lowest variance ① use all info ② remove outlier
③ Consistency $\text{plim}_{n \to \infty} \bar{\beta} \to \beta$

① check linearity: polynom. terms, correl
② interaction
③ Collinearity
④ heteroscediasticity
⑤ Outliers

Outlier:
① Remove top 5%
② bottom 5%

F-val: not sign → pooled
sign = Reject → setterwait

t-test: Categorical vs. quantitative, but two levels
chi-sq: Categorical vars could have multiple levels

t-test Comparison of two groups (age of Democrats vs. Republics)
Chi-sq test Comparison of more than two groups (from max likelihood) ⇒ ethnic group vs. presidential choice

```
Data ...;
infile ...
input ...
run;

proc syslin ...2SLS;
endogen ...;
instrument ...;
4xmodel ...;
run

Proc corr;

Proc freq;
Tables x*y Chisq;
run;

proc cluster method=centroid out=...;
id ...;
Var ...
run;

proc fastclus ... maxclusters...;
id ...;
Var ...
run;

proc model outtest=...
eq.first= ...;
eq.second= ...;
solve ... /solve print;
id;
run;

Put ...(keep...)
if ... then delete;

proc sql;
creat table ... as
select ..., ...
from ..., ...
wher ... = ...;
run

proc discrim ...;
class ...;
Var ...;
run;

proc sort ...
by ...

Data ...; obs=1;
set ...;
array ... 1-s1c;
select(...);
when("s") do;
...
end;
otherwise ...;
end;
```

Heckman sample selection

- Two latent variables

$Y_1 = \beta' X + U_1$ $\quad$ $Corr(u_1, u_2) = \rho$

$Y_2 = \gamma' Z + U_2$: selection equation — Y_2: latent var: propensity of being included in the sample

Y_1 observed if $Y_2 > 0$ $\quad \begin{cases} Y = Y_1 & Y_2 > 0 \\ \text{missing} & Y_2 \leq 0 \end{cases}$

Bias: $E(Y_1 | Y_2 > 0, X, Z) = \beta' X + \rho\sigma \dfrac{f(\gamma' Z)}{F(\gamma' Z)}$ $\quad$ σ^2: Variance of U_1

$E[Y_1 | Y_2 > 0, X] = \beta' X + \rho\sigma E[f(\gamma' Z)/F(\gamma' Z) | X]$

Expectation over Z

⇒ max likelihood estimation

$h(y | X, Z, \beta, \gamma, \rho, \sigma) = [\sigma F(\gamma' Z)]^{-1} f((y - \beta' X)/\sigma) F[(\rho \frac{(y-\beta' X)}{\sigma} + \gamma' Z] / \sqrt{1-\rho^2}$

Conditional density
= write things according to others

D_j Dummy of missing $\begin{cases} D_j = 0 & \text{if } Y_j \text{ is missing value} \\ D_j = 1 & \text{if not} \end{cases}$

$\ln L(\beta, \gamma, \rho, \sigma) = \sum_j \{ D_j \cdot \ln[h(Y_j | X_j, Z_j, \beta, \gamma, \rho, \sigma)]$
$+ D_j \cdot \ln[F(\gamma' Z_j)] + (1 - D_j) \cdot \ln[\cancel{F(\gamma' Z_j)}\ 1 - F(\gamma' Z_j)]$

$y_i = x_i \beta + u$

$z_i^* = w_i' \alpha + e_i$

$z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$ $\quad Pr(z_i = 1) = \Phi(\alpha' w_i)$ selection model — probit

$\ln L = \sum_{z=0} \ln(1 - \Phi(w_i \alpha)) + \sum_{z=1} \ln\left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right) + \sum_{z=1} \frac{1}{2\sigma_u^2}(y_i - x_i' \beta)^2$

$+ \sum_{z=1} \ln \Phi\left(\dfrac{w_i \alpha + \rho\left(\frac{y_i - x_i' \beta}{\sigma_u}\right)}{\sqrt{(1-\rho)^2}} \right)$

Latent class model (LCM)

- latent var: discrete : chance of having one value
- Finding subtypes: latent classes
- measure uncorrelated with each distribution → latent profile analysis
- assign social class (on ability or merit) → example
- observed variables usually discrete
- within each latent class variables are statistically independent (helps to disentangle dependence)

$P_{i_1, i_2, \ldots, i_N} \approx \sum_t^T p_t \prod_n^N p_{i_n, t}^n$ → marginal or conditional probabilities

p_t → recruitment or unconditional probabilities, sum to one

T: # latent class

- political questionnaire ⇒ latent variable: political opinion
  latent class: political groups

P(chance of certain answer chosen | Given group membership)

- constraints the classes to form a segments of a single dimension (factor analysis)

Multivariate mixture estimation (MME)

- applicable to continuous data
- data arise from mixture distribution (e.g. mixture of men & women)
  ↳ e.g. (whether unobserved; but action observed)

$L(\beta_{1:k}, \pi; Data) = \sum_{d=1}^{D} \log \sum_{z=1}^{k} p(z | \pi) \prod_{n=1}^{N} P(w_{d,n} | \beta_z)$

Data analysis Final Exam items

Focus on (1) Codes & (2) interpretation

1. Tobit, model selection, ..., Likelihood
   logit, nested logit, probit
2. Survival analysis, Phreg, Lifetest, Lifereg
3. Gauss mixture, latent $\Longrightarrow$ HW codes

$R^2 \longrightarrow$ which model is better

Log likelihood $\rightarrow$ only relevant in nested models (e.g intercept only model)

---

logit $P_1 = \frac{e^{V_1}}{e^{V_1}+e^{V_2}} = \frac{1}{1+e^{V_2-V_1}}$

For non-choice specific $\Rightarrow$ two seperate intercept, but keep one to identify (different Coeff for diffrnt brands)

$\Rightarrow$ Reference brand (like interaction of brand & income)

effect of income depend on which brand you purchase

$\Rightarrow$ diffrnt elasticity of income w.r.t choice

---

mixture model $\longrightarrow$ nested for ~~normal~~ model

take away: $-\log(1) = 0$

---

print val. Gaus: (1) p1;; B1;; F;; L

(2) Variance $\rightarrow$ positive $\longrightarrow$ why SSqrt(Σ), $\Sigma = s \times s$ (positive definite or positive semi definite)

$\Sigma = C'C$ cholesky decomposition

(3) transformation $\beta = \frac{e^a}{1+e^a}$ $\beta \in (0,1)$

Gaussian Quadrature

$v_i = a_i + b_i I_i + \epsilon_i$

$a_i = \bar{a}_i + \epsilon_i \sim N(0, \sigma_{a_i}^2)$ random & not fixed

$P(i|a_i) = \frac{e^{V_i}}{1+\sum e^{V_i}}$

$P(i) = \int_{-\infty}^{\infty} P(i|a_i)\phi(a_i)\,da_i$ trivariate normal

addition easy in computer compared to integral

$\int_{-1}^{1} f(x)dx = \sum w(x) g(x)$ vs Trapezoid rule / Gauss Hermit (area calc by two points)

$\hookrightarrow P(i|a_i)$ — weights

reason: need integration (at household level) {intercept, Coeff

discrete hetrogenity: $(a_1,p_1)\ (a_2,p_2)\ (a_3,p_3)$

bimodal: two var with normal dist. (a_1, a_2, p_1, p_2)

Continuous: more than two vars & they are correlated (Var-Cov Matrix)

DEA and SFA (1) Data envelop analysis (2) Stochastic frontier analysis

$\hookrightarrow$ Production function estimation

SFA: noisy frontier $y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

multiple outputs, multiple inputs, not stochastic

- measure efficiency — capture the Best vs. OLS: AVG $\hookrightarrow$ max output given input

DEA: Capture efficient frontier relation b/w controlable input and output $\longrightarrow$ Linear Programming

observable efficiency: $V_0 = \frac{w_1 O_1 + w_2 O_2 + w_3 O_3}{v_1 I_1 + v_2 I_2 + \dots + v_n I_n}$ $\longrightarrow$ optimal weights

$(n-1)$ Constraints — Cross section data

---

1. - move line $\rightarrow$ every points at its right

Slack: horizontal or vertical distance from optimal level (2)

- Constant return to scale assumption (CCR)

BCC model: allows varying return to scale

pareto optimal $\equiv$ production function frontier

SFA: $\varepsilon = u + v$

u: random error term; v: one sided Error term of efficiency (truncated normal, exponential, or gamma)

Error is in SFA, but in DEA there is no Error

PCA & Factor Analysis: summerize outputs in the index

Dummy Variable: Comparison with option left out

Weibull: generalization of Exponential distribution

Model fit: Compares model with intercept with model with Covariate

"." in SAS: model does not scale one model so huge that #'s not close

Sol: divide by 1000

Life test: Compare survival of two or more groups $\hookrightarrow$ univariate analysis

MLE is used

Competing risk specification: more than one hazard (e.g smoke, obes...) independent Causes, Compete for same reason

- for studying time before event (transition from state)

- problem { (1) not normal dist (2) Censoring $\Rightarrow$ Reg will give bias estimate

Logit $\rightarrow$ ignores time info $\Rightarrow$ inefficient

LifeReg: not allow time dependent Covariates (Continuous)

PHReg: Proportional hazard model, Cox partial likelihood (Continuous & discrete time data), not requir dist. assump

✓ time dependent Covariate

dependent Var: hazard rate; died next instance given not died until now

Right/left/Interval Censoring

Type 1: under Control of investigator (e.g 3 year)

Type 2: prespecified # of events occured (e.g 50/100 died)

Random Censoring: not under Control (e.g participant leaves) or random entry

Time is continuous

$\hookrightarrow$ problem of Bias

Survivor function: $S(t) = 1 - F(t) = Pr(T > t)$

Hazard function: $h(t) = \lim_{\Delta t \to 0} Pr(t \le T < t + \Delta t)$

$S(t) = e^{-\int h(u)du}$

$h(t) = \frac{f(t)}{1-F(t)}$

lower bound zero but $\neq$ Probability

$\frac{1}{h(t)} \rightarrow$ time b/w another event

assum. Constant hazard rate

$\Rightarrow \log h(t) = a$, $S(t) = e^{-at}$, $f(t) = ae^{-at}$

$\log h(t) = a + bt \Rightarrow$ time has Gompertz dist.

$\log h(t) = a + b\log(t) \Rightarrow$ weibull dist

```
proc Lifetest;
time dur*status(0);
Strata ...;
run
```

```
proc lifereg;
Model week*arrest(0)=...
/dist=lnormal;
Output out= ;
run;
```

PROC lifereg

$\underbrace{\log T_i}$ $= \beta X_i + \sigma \varepsilon_i$ → ensures predicted values of T are positive

✓ right & left censoring

✓ Hypothesis of shape of hazard

✓ shape known ⇒ efficient estimate

$S_i(t) = S_j(\Phi_{ij}t)$ → rate of age

* e^{β} → if X binary: ratio of expected survival time

AFT: accelerated failure time model

* Continuous $100(e^{\beta}-1)$: % increase in expected survival time for each unit increase in the var.

✓ automatic creation of Dummy

③

PHREG — David Cox — no assump of dist ⇒ semiparametric

✓ allows only right censoring

✓ hazard ratio / risk ratio: e^{β}

- no intercept in chow. of likeli.

✓ $100(e^{\beta}-1)$ ⇒ % change in hazard of arrest

```
Proc phreg data=...;
model week*arrest(0) = ...;
run;
```

Continuous variable interpretation — how much increase or decrease

Logit & probit interpretation
- think in terms of elasticities
- Convert betas to elasticities

Mixture & latent model elasticities
- put values & check response — do numerical integration
e.g. multiply in 1.05 for 5% change & check change in prob
⇒ average change

IIA — independent from irrelevant alternative: the probability of making choice only depend on their own characteristics and not the third choices characteristics

likelihood $= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$ $\quad f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(y-x\beta)^2}$

$\Rightarrow \sum x_i \log(\theta) + (1-x_i)\log(1-\theta)$ $\quad \sum\left[-\frac{1}{2\sigma^2}(Y-X\beta)^2\right] - \frac{n}{2}\ln(2\pi)$

$LL = -\frac{n}{2}\ln(\sigma^2) + \sum\left[-\frac{1}{2\sigma^2}(Y-X\beta)^2\right] - \frac{n}{2}\ln(2\pi)$

Mixture model
- person can be in state 1 or 2: $P, (1-P)$
- brand fixed effect → intrinsic pref.

$\underset{\text{choice 1}}{P(1)} = q \cdot P(\text{choice}=1 \mid \text{state}=1) + (1-q)P(\text{choice}=1 \mid \text{state}=2)$

$q = f(\text{income, weekday, weekend})$ ⇒ logistic — person decides to be searcher or non searcher

Brand intercepts should be $n-1$ → since work with dummies compared with reference

3 state { Calculate all brands: q; Evaluate favorite brand: r; No evaluation: $1-q-r$ }

latent model — segmentation { first segment: s; second segment: $1-s$ }

$L_i = S(A_1) + (1-S)(A_2)$, $L = \prod_{i=1}^{N} L_i$

- different with cluster analysis: segment based on unobserved parameters (latent)

④

- Discriminant analysis assumption: multivariate normal
- when discrete var. ⇒ better to use: logit
- t-test: assumes vars' are independent, does not control for other variables
- Heterogeneity & endogeneity role: give bias to estimates ⇒ make t-test estimate biased
- Factor analysis: not data reduction (PCA), but to find underlying constructs
- # Excluded var. from equation $\boxed{\geq}$ greater than # endogeneous var minus one
- Fix one = intercept effect & not coeff effect
- Do data analysis when missing data < proc freq / proc mean
- check size of segment even when significant
- $\frac{\beta_{logit}}{\beta_{probit}} = 1.70$

Gauss
- Pico prog.txt // edit
- Pwd // where in system
- Proc mdc: multinom logit

$\begin{cases} y_1 = a_1 + b_1 y_2 + c_1 x_1 \\ y_2 = a_2 + b_2 y_2 + c_2 x_1 + d_1 x_2 \end{cases}$ $\quad x_1/y_1 = \frac{(c_1 + c_2 b_1)}{1 - b_1 b_2}$

$AIC = -2\ln(L) + 2k$ ⇒ lower better → penalization of # of variables (Akaike info criteria)

$SC = -2\ln(L) + \ln(n) \cdot k$

SBC: shwartz bayesian criterion

BIC: bayesian information criterion

$R^2 = 1 - \frac{L}{L_0}$ — L → full model; L_0 → null model with no covar (intercept only)

{Bus, Car, Train} ⇒ Nominal var.

{A, B, C, D} ⇒ ordinal var. ⇒ cut off point for classification

$\frac{\partial pr_1}{\partial P_1} \cdot \frac{P_1}{pr_1} = \beta(1-pr_1) \cdot Pr_1$: own elasticity

$\frac{\partial pr_1}{\partial P_2} \cdot \frac{P_2}{pr_1} = \beta \cdot pr_1 \cdot P_2$: cross price elasticity

Proc MDC → ~~Conditional logit~~

Solution to IIA
① Nested logit model: Error term: GEV, generalized extreme value
② Random coeff logit → GEV — random effect in logit model
③ multivariate probit: multivariate normal — GHK estimator: simulation based

Correl → common factor affect both error term & first and second choice

Jobt

- lots of Zeros $\Rightarrow$ - Censored regression
  $\begin{cases} y_i = y_i^* & y_i^* > 0 \\ c & \text{otherwise} \end{cases}$ (Zero means have not crossed the threshold)

- truncation: only observed values $> c$
  make area under the curve equal to one

$f(y^* | y^* > 0) = \frac{c-\mu}{\sigma}$ — limited/Censored dependent variable

$\prod_{y_i^* > c} \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \prod_{y_i^* \le c} \Phi\left(\frac{c-\mu}{\sigma}\right)$

```
proc QLIM ..;
  model ... = .... /discrete;
  model ..... = .../ select(active=1);
run;
```

```
proc qlim data=...;
  model profit=...;
  endogeneous profit ~ Censored (lb=...  ub=...);
run;
```

```
proc logistic data=... descend;
  class affinity ....;   (tells discrete)
  model active = ... /link probit;
run
```

```
proc mdc data=...;
  model decision ... /type clogit
    nchoice=4;
  id i;
run;
```

Gauss multinom logit

```
do ....     ∀ all i < 3129
  u1 = e ....;
  u2 = e ....;
  pen = u1 + u2 + ...;
  if (choice=i) LOGU = ui/pen; endif;
  like = like + ln(LOGU);
endo
ret(-lik)
endp
```

Gauss mixture

```
new LNS <-> logv x LNS
    LS  <-  Logv x LS
it find household or
the next diff:
Lik = Lik + LN(q x LS +
(1-q) x LNS);
```

Gauss general: Extract Coeff (params) first
then extract values per i (e.g. household)
third Calculate utility for each choice

- Clustering: identity behavior $\Rightarrow$ Discriminant $\rightarrow$ Connect to demographic
  $\hookrightarrow$ look to mean to name clusters

- Variance much larger than mean $\Rightarrow$ negative binomial

- large proportion of Zeros $\Rightarrow$ Zero inflated poisson

---

① log & not mult. likelihood $\Rightarrow$ due to small # Computer capability

$\boxed{SUR} + \boxed{2SLS} \Rightarrow \boxed{3SLS}$

② minimizing least square or weighted least square

③ GMM (General method of moment) $\rightarrow$ minimize least square with weights equal to Covariates

---

Problem of Regression on discrete:
① Hetroscedasticity
② prediction

$\begin{cases} U_1 = \alpha + \beta X_1 + \varepsilon_1 \\ U_2 = \alpha_2 + \beta X_2 + \varepsilon_2 \end{cases}$ $P(\text{choice}=1) = P(U_1 > U_2) = P(\alpha + \beta X_1 + \varepsilon_1 > \alpha_2 + \beta X_2 + \varepsilon_2) = P(\varepsilon_1 - \varepsilon_2 > \alpha_2 - \alpha + \beta(X_2 - X_1))$

odds $\leftarrow$ $\log\left(\frac{P}{1-P}\right) = \alpha + \beta X$ linear log of odds

$\begin{cases} U_1 = V_1 + \varepsilon_1 \\ U_2 = V_2 + \varepsilon_2 \end{cases}$

logistic

$\begin{cases} Y_i^{0*} = \beta_0 X_i + \varepsilon_0 \\ Y_i^{1*} = \beta_1 X_i + \varepsilon_1 \end{cases}$

$Y_i \begin{cases} 1 & \text{if } Y_i^{1*} > Y_i^{0*} \\ 0 & \text{Otherwise} \end{cases}$

$\begin{cases} \text{Gumbel: } P(x) = \frac{1}{\beta} e^{\frac{x-\alpha}{\beta}} - e^{\frac{x-\alpha}{\beta}} \\ \text{inverse Gumbel: } \alpha - \beta \ln\left(\ln\left(\frac{1}{\beta}\right)\right) \end{cases}$

$U_{ni} = \beta Z_{ni} + \varepsilon_{ni}$

$P_{ni} = \text{Prob}(Y_{ni} = 1) = \text{prob}(U_{ni} > U_{nj} \ \forall j \ne i) = \text{Prob}(\beta Z_{ni} + \varepsilon_{ni} > \beta Z_{nj} + \varepsilon_{nj}, \forall j \ne i) = \text{Prob}(\varepsilon_{nj} - \varepsilon_{ni} < \beta Z_{ni} - \beta Z$

$= \frac{e^{\beta Z_{n1}}}{e^{\beta Z_{n1}} + e^{\beta Z_{n2}}}$

weighted logistic regression:

$\beta = (X^T X)^{-1} X^T Y$ $\quad Var(X) = n \times p \times (1-p)$

$\beta = (X^T W X)^{-1} X^T W U$ $\quad W_{ii} = \mu_i (1 - \mu_i)$

$U_i = X_i \beta + \frac{y_i - \mu_i}{W_{ii}}$

$W_{ii} = \mu_i \times (1 - \mu_i)$

$\mu_i = \frac{1}{1 + e^{-X_i \beta}}$

$X \sim Bin(P, N)$

SAS Prof Murthi

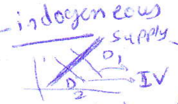
→ diff ppl diff Price sensitivity

- unobserved hetrogeneity → fix effect — 1st problem Heterogeneity
- fix effect in Coeff $b = \gamma + \delta\,\mathrm{Inc}\,age + age^2$
- Observed hetro ↳ = hierarchical model ↳ interaction
- average probability → predicted market share
- Fixed effect → 99 dummies α_t
- Hauseman test → random effect is Correct?
  Null hyp: Random effect Correct Model

PROC TSCSREG
① id ③ CS ② time
→ sort data
proc panel
proc mixed

Simultaneous system of Equations — X ⇄ Y dependance

OLS: exogen assump. $\begin{cases} y = f(x) \\ x = g(y) \end{cases}$ — indogeneous



- identification problem — identified
  # variable < # equations
  identified = ✓ → over identified — multiple solutions
- high correlate price, less correlate with sales — 2nd problem Endogeneity
  affect sales through price — biased estimate — proc syslin
  ~~based~~ Correl. only if stoch & not fixed
- Error term (ε) → part of y not Explained by x (mean)
  ↳ mean variation of stochastic y

if $E(\varepsilon) \neq 0$ ⇒ Biased $\hat{y} = X\beta$ ⇒ q-q plot to check Normality

Assump: Error ~ Normal → OLS
  if not normal → non parametric reg
  not normal (theory) ⇒ log normal as option

+ Over identified → multiple solution ⇒ 2SLS
  excluded Predetermind var > # included indogeneous −1 minus one exogeneous
  (e.g 2-1)

2SLS: Stage 1 — indogen var regress among all predetermined P/y,t
  ⇒ $\hat{P}_t$: indep Error term — or start with $\hat{q}$-$\hat{q}$
  Stage 2 — use it in the first equation $q/\hat{P}$

⇒ no oversupply of instrument

Haeseman test of simultaneity by check Correlation
$y_1 = a_1 + b_1 y_2 + c_1 x$
$y_2 = a_2 + b_2 y_1 + c_2 x$
⇒ $y_1/x_1 = \frac{c_1 + c_2 b_1}{1 - b_1 b_2}$

SUR — Seemingly unrelated Regression → Pooling — Stack data

$\mathrm{Cov}(\varepsilon_1, \varepsilon_2) \neq 0$ — e.g Coke & pepsi sales → Not b/c of simut. but diff process

Efficiency → variance

Proc Syslin
R-sq ≠ % Variance

simultaneity ⊕ SUR ⇒ 3SLS
$\mathrm{Corr}(\varepsilon_1, \varepsilon_2)$

→ Simultaniety
{ LMLE (Limited info)
{ FIML (Full information) → 3SLS

```
%let dsn=clinics; replace word    use &dsn
% macro look;    use %look(clinics, 10)
%MEnd look;    %Do i=1 %To 5;
```

why endogeneity
① simultaneity
② ~~Simultaneity~~ Error in measurement
③ ommited variable

Factor analysis — Solution to multicollinearity

measure-index: summary of multiple factors (condense) Composite

information = variation $\omega_1^2 + \omega_2^2 + \dots + \omega_n^2 = 1$

PCA: Principle Factor Analysis = weight?

③ ↳ ensures no correlation (orthogonal)

① Principal Component → most variance { angle max var — Diminishing return

② squares of weights add up one ⇒ scale — eigen vector

PCA	Factor analysis
no Error term	Error term
Data reduction	Maybe not correl data reduction
uncorrel	
most var	understand underlying factor construct
Composit indx	

⊕ Factor loading: weight of each factor for each of the variables sign ≠ impt
⊕ loading more than .5 since $a^2 > \ldots$
- Calculate predicted value (per person)

2 ⊕ Communality → factors all combined how capture variance

⊕ percentage of variation → how much of variance explained by each of factors

3 ↳ squar of factor loading ⇒ eigen value ⇒ % variation explained

% variance Explained by the factor → $\frac{\text{eigen}}{n}$

Before factor analysis: either mean correction ‖ Standardization
deduct mean from original values — $\mathrm{Var} = 1$ $\mu = 0$

4 Factor score: Predicted value of factors — positioning map

5 SSCP: Sum of square cross product $x_1^* = \cos\theta \cdot x_1 + \sin\theta \cdot x_2$

→ Factor rotation: method, better intrprt, new constraint
→ Varimax rotation: get loading close to zero or one
→ select ① at least ⊗ var > 1/f → # factors
  ② Eigen value > 1 → elbow (diminishing return)
- Annotated factor analysis

Indogeneity
① Latent Instrumental Variable (LIV) → max likelihood
② Propensity score matching → score - mk role of other var insignificant
  ↳ Heckman

Cluster analysis — group of similar objects

Factor analy	Cluster an
Group similar vars.	Group similar Obs

techniques ① Decision tree
② AID: Automatic interaction detection
③ CHAID: chi-sq AID

① euclidian distance ② block distance ③ mahalanobis distance stat distance when factors are correlated

Distance b/w groups: ① min distance: single linkage
② max distance ③ AVG distance: AVG linkage
④ Distance b/w controids

Output: ① how many groups to keep ② which group each element blong to

type method: ① hierarchical - dendogram - tree - cut < 50 Obs — For Exploratory
② non-hierarchical prespecify # cluster - iteration

RMSTD: large SPR: small R^2: close to 1 — Distance b/w two in cluster: small

Cluster → Based on behavioral variables & not exogeneous ≠ age/income
e.g profit: interest of firm

Factor analysis on behavior → Cluster analysis on behavioral var → Discriminant/logit analysis: demographic media habit → Targetting → Distinguished charactertics

⊗ → output of SAS → Tex → use excel in middle mixture model

$P(i \mid \text{search}) = \dfrac{e^{u_i}}{\sum e^{u_k}}$ $\quad L(i \mid \text{search mode}) = \prod \dfrac{e^{u_i}}{\sum e^{u_k}} = \prod (i \mid \text{search mode})$

$P(i \mid \text{non search}) = \dfrac{e^{v_i}}{\sum e^{v_k}}$ $\quad L(i \mid \text{non search mode}) = \prod \dfrac{e^{v_i}}{\sum e^{v_k}} = \prod p(i \mid \text{non search mode})$

$L(i) = q \times L(i \mid \text{search mode}) + (1-q) \times L(i \mid \text{non search mode})$

- unobserved hetrogeneity
Cluster according to unobserved

$LL = \sum_i \ln(L(i))$

$q = \dfrac{e^{a + b\,\text{Inc} + c\,\text{FS}}}{1 + e^{a + b\,\text{In} + c\,\text{FS}}}$ → probability of being in one state

---

Survival analysis

Experiment stop after 10 die ← Type 2 → manual termination

observer Control: Type 1 Censoring

Left/Right/interval censoring

- time between events
- state transition (marriage)
- Factors affect time to transition
- Censoring problem ⇒ Regression biased estimate
- logit : whether switch or not - survival: time to switch
- Univariate analysis: lifereg

problem of: ① Normal error Reg ② Censoring

life reg: $\log T_i = \beta X_i + \sigma \varepsilon_i$ not allow time dependent Covariate

- PHREG: Cox's partial likelihood method semiparametric no dist assumptions - hazard/risk ratio: $\exp(\beta)$

% Δ hazard: $100(e^{\beta}-1)$

proportional hazard

dependent var = hazard rate = happen next instance given not happen until now

- Random censoring → terminate for reasons not under Control of researcher = leave study

$S(t) = 1 - F(t)$ → Survivor func

$h(t) = \dfrac{f(t)}{S(t)}$ hazard func

accelerated failure time model (AFT) → life Reg

↳ Not probability but number greater than one

$\dfrac{1}{h(t)}$ → Time betwen another event → given Constant

Insurance

$\log h(t) = a \Rightarrow S(t) = e^{-at} \Rightarrow f(t) = a e^{-at}$

Cox: $\dfrac{h(t \mid X_F)}{h(t \mid X_Z)} = \dfrac{e^{x_i \beta}}{e^{x_Z \beta}}$

$\log(h(t)) = a + bt$ → Gompertz distribution

$\log(h(t)) = a + b \log(t)$ → weibull distribution

life test → kaplan meier (km) for small data sets
life table method : large data sets

$\exp(\beta)$ → ratio of Expected Survival b/w groups
X : binary

$100(\exp(\beta) - 1)$ → % increase expected survival time each unit increase in var

---

- IIA: independent from irrelevant alternative: prob of making choice only depends on their own characteristics and not on third choice {A, B} adding C does not change preference

---

$Lik = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} \rightarrow LL = \sum x_i \ln(\theta) + (1 - x_i) \ln(1-\theta)$

$LL = -\frac{n}{2} \ln(\sigma^2) + \sum \left[-\frac{(Y - X\beta)^2}{2\sigma^2} \right] - \frac{n}{2} \ln(2\pi)$

Discriminant → assumption of multivariate Normal

Discrete variable → logit is better

ttest → assumption of independance of vars

Hetrogeneity & endogeneity → biased estimate

Factor analysis → goal: find underlying Constructs not measurable on one dimension

Excluded var from equation > # endogenous var - 1

Fix one → intercept effect & not Coefficient effect

proc freq, proc mean → analysis of high level of data (missing)

$\dfrac{\beta_{logit}}{\beta_{probit}} = 1.70$

Bass model

$Q_t = p(\bar{Q} - N_t) + r\left(\dfrac{N_t}{\bar{Q}}\right)(\bar{Q} - N_t) = \left(p + r\dfrac{N_t}{\bar{Q}}\right)(\bar{Q} - N_t)$

(r → q, $\bar{Q}$ → M)

assumption: MKT potential remains Constant over time
⇒ max likelihood works better

- Commonality $= a_{11}^2 + a_{12}^2$ $\quad x_1 = a_{11} F_1 + a_{12} F_2 + \cdots$
- eigenvalue = sum of sq of loading each factor $\equiv$ % var × N

e.g. $\lambda_1 = r_{11}^2 + r_{12}^2 + \cdots + r_{1N}^2$

$F_1 = r_{11} x_1 + r_{12} x_2 + \cdots + r_{1N} x_N$ $\quad r_{ij}$ → Factor loading

- Percent of variance: $\dfrac{a_{11}^2 + a_{21}^2}{2}$ For each factor

SAS prof Murthi — Discriminant analysis

- Seperate groups [high] cutting score classify good & Bad
- Find discrim. func, $\lambda = \frac{SS_b}{SS_w}$ → Σb_i^2 → between group; SS_w → $\Sigma \omega_i^2$ → withing group %
  - line seperate member of two groups — 3 line four group
- Classific func → cluster prediction — { Coupling ↓ / Cohesiveness ↑ }

method of classific:
1. Cut off value method
2. Decision theory approach
3. classific. func. approach
4. Mahalanobis Distance method

OLS when Discrete: ① high heterodescediasticity or Bias ② prediction will lie b/w zero & one and beyond of zero & one

1. Discrim analys OR
2. Logit → no assumption

↳ x ~ distribute multivariate normal (MVN) — significant diff mean b/w groups

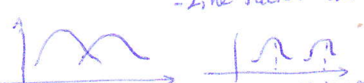
e.g dependent var could be discrete

- Discrim analysis output → weights — classify based on — * Score according to two scoring functi (module covered) → assign to group with higher cohesiveness → e.g high buyer Java module / Low buyer Java module

- Poisson Regression
  1. Counts → not normal distributions
  - Condition to use: same mean & Variance
  - Negative binomial dist — when Variance is much Greater than the mean
  - Zero inflated regression — when # zeros '0' are high — two segments: one with zero, one with positive
  - proc genmod

- Discriminant analysis ① Variable summerize persimoneously
(DA) different b/w groups ② classify future obs into groups
  - discriminator: variable that give best discrimination — like factor analysis (FA)

Homogenity of groups


less H. — more homogen

FA → max Var = max SS_t — DA → max b/w group to withingroup sum of squares $\frac{SS_b}{SS_w}$

classification → using cut off value on calculated scores

- differen b/w two groups using t-test b/w variables (υ_i)
- $Z = \Sigma w_i \upsilon_i$ * been classified before

Reporting: when you talk about sth, show it in the table

Clustering (identify behavior) → connect to discrim var (demographic)
↓ name of groups

Likelihood: $P(y|x)$ if independent vr: multiplications of them
MLE: problem: local maximum: run multiple seeds

- tailor expesion $D \ln L(\tilde\theta) \simeq D \ln L(\theta) + (\tilde\theta - \theta) D^2 \ln L(\theta)$
- Regularity Condition: info matrix converge $1/T$
- $(0, I(\theta))$ → $D\ln L(\theta) = \Sigma^T \frac{\partial}{\partial\theta}\ln f(y_t|\theta) = \Sigma^T D\ln f(y_t|\theta)$
- Cramer-Rao $\xi = \frac{-D\ln L(\theta)}{\sqrt{T}\sqrt{\frac{I(\theta)}{T}}} \xrightarrow{asy} N(0,1)$
- $\sqrt{T}(\tilde\theta - \theta) \rightarrow \frac{\sqrt{\frac{I(\theta)}{T}}}{\frac{D^2\ln L(\theta)}{T}} \xi \Rightarrow \sqrt{T}(\tilde\theta-\theta) \xrightarrow{asy} N(0, IA(\theta)^{-1})$

GMM: minimizing least square with weight equal to cours: e.g: $Cov(\varepsilon_i, \theta)$, $Cov(\varepsilon_1, \varepsilon_2)$ → diff objective func
better for: ① auto correlation / serial corr

$x \sim MVN$ = Correl ⇒ discrim ($Y \sim$ discrete)
$x \sim$ discrete ⇒ logit $P(1)$ $P(2) = 1 - P(1)$

OLS problem: ① Heterodescedasticity ② prediction — Gumbel

V_1: $U_1 = \alpha_1 + \beta x_1 + \varepsilon_1$, $U_2 = \alpha_2 + \beta x_2 + \varepsilon_2$ — $\varepsilon_1, \varepsilon_2 \sim$ extreme value

$P(1) = P(U_1 > U_2) = P(\alpha_1 + \beta x_1 + \varepsilon_1 > \alpha_2 + \beta x_2 + \varepsilon_2) = P(\varepsilon_1 - \varepsilon_2 > \alpha_2 - \alpha_1 + \beta(x_2 - x_1))$

→ odd = chance win against losing

$P = \frac{1}{1+e^{\alpha + \beta x}}$ $\ln(\frac{P}{1-P}) = \alpha + \beta x$ (coding) → ln odd ~ linear $P = \frac{O}{O+1}$

proc logistic / class discrete / descending — Compare intercept only

AIC: lower better — Akaike info criteria
SC: shwartz criteria (lower better) = BIC: Bayesian info criteria
LRT: likelihood ratio test $= \frac{L_0}{L_1}$ — $SC = -2\ln(L) + \ln(n)K$
$AIC = 2K - 2\ln(L)$ — SBC → Bayesian

interpretation: based on odds — $R^2 = 1 - \frac{LL}{LL_0}$ → intercept only

$P(1) = \frac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3}} = \frac{1}{1 + e^{V_2 - V_1} + e^{V_3 - V_1}}$ — ① x ~ choice invarient → if x person specific ②

③ one intercept since only difference identified — $V_1 = \alpha_1 + \beta_1 Ind_1$, $V_2 = \alpha + \beta_2 Ind_2$: seperate beta ⇐ Cancel out

Conditional logit = Multinomial logit — proc MDC ↳ code dummy var

IIA: independence from irrelevent alternatives — same cross elasticity

not independent: ① multinomial prob.t → GHK simulation → multivariate normal
Generalize extreme value (GEV) ← ② nested logit (tree): branch — category homogen → correl b/w brands
Random effects logit model — ③ Random Coefficient $\beta_i = \beta + E$ (RCL) Random coeff. logit (BLP) → Berry Lebinson Pakes 95 — ④ FCMNL: Flexible multinomial logit

ordinal → cut off points — $Pr(1) = \frac{e^{a_1 + bP_1}}{e^{a_1 + bP_1} + e^{bP_2}}$

own price elasticity $= \frac{\partial Pr_1}{\partial P_1} \cdot \frac{P_1}{Pr_1} = b(1 - Pr_1) \cdot P_1$

Cross elasticity $= \frac{\partial Pr_1}{\partial P_2} \cdot \frac{P_2}{Pr_1} = bPr_1 \cdot P_2$

Correlation = Common factor — $\beta_{probit} = k\beta_{logit}$

Tobit ① indep: discrete / continuous ② Lots of zeros
→ Censored reg: below threshold → observe zero

Truncation: only observe values lower/greater than c $\{y^* \ y^* > c$ otherwise
↳ try to make area under curve of dist = one

$\prod_{y_i^* > c} \frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma}) \prod_{y_i^* \le c} \Phi(\frac{c-\mu}{\sigma})$ — limited dependent var — Censored

Proc GLIM — gradient search method ↳ also ordered logit — Hetresced could exist

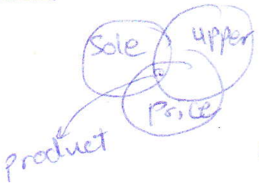
① elasticity $\frac{\Delta D}{\Delta P} \cdot \frac{P}{D}$ → Coefficient
② one unit change = Coeff ③ effect: mean × Coeff
③ interactive effect (mean param × coeff)

- whole idea of likelihood, mixture, random coeff...: define objective function (e.g error) and minimize or maximize it — tobit

```
proc qlim data=;
  model y=x1;
  endogeneous y~discrete;
run
```

```
proc qlim ...;  meth=qn
  model ... /discrete (d=logit);
  hetero int ~ .../noconst;
  endogeneous yn, censored (lb=0)
run;
```

$\boxed{\text{Conjoint}}$ Product = Combination of attributes



① Valuation of each attrib ?
② importance of attrib ?

$u_{p1} + u_{p01} + u_r = 0$

- Ranking
- not all Comb. but Samples
- Preference : inverted Rank
- Run Regression on each person
- Part worth value of each level
- utilities are additive (assumption)
- Dummy Regression

- u_{min} & u_{max} importance rate $I = (Max\,U - Min\,U)$ → Variation & Difference
- Relative importance : Divide by sum

$$w_i = \frac{I_i}{\sum_{i=1}^{m} I_i}$$

$k_i - 1$ Dummy var (levels of attrib i)

- maximum utility rule : product most likely to purchase [problem when very close utilities]
- probab. of purchase for each : Luce Rule $A = U_A / \sum_{i=1}^{N} U_i \Rightarrow$ Expected mkT Share
  → problem: either positive or negative probabilities (Neg prob) - for discrete preference
- logit rule: $p(A) = \dfrac{e^{U_A}}{\sum_{i=1}^{N} e^{U_i}}$ - Segmentation, what if, NPD
- ODS

$$\begin{cases} \underbrace{u_N - u_L}_{\text{part worth}} = b_1 \\ u_C - u_L = b_2 \\ u_C + u_N + u_L = 0 \end{cases}$$

Gaus

- Scale variables to make them managable/1000

proc likFun (Beta)

loop

LIK = lik + LN (logv); → prob of choice (x) based on whic one is selected & its spec

Retp (-lIK);

ENDP;

x_0 = {1, 1, 2, -1} → fixed effect brands → price coeff

{beta, lik, G, retCode} = sqpsolve (&LIKFUN, x0);

H = hessp (&likfun, para.);

Cov = inv(h);

stderr = sqrt(Diag(Cov)); Trat = Para./std err;

output ON; para; trat; STDERR; MINI; OUTPUT OFF;

pwd → where you are in system

pico → edit

$\boxed{\text{Std err} = \text{sqrt}(\text{inv}(\text{hessian}))}$

e.g $\frac{\partial \text{ variation of } x_1}{\partial x_2} = \frac{\partial f}{\partial x_1 \partial x_2}$

= how two changes are correlated = Cov