Prof. H. D. Vinod, class notes for Econometrics II students, , April 20, 1999.
Ch. 20 of 3rd edition of Greene pp 950 onward:
THM 20.1 If x has pdf of f(x), a is a constant, so keep all stuff to the right of a (x>a)

f(x|x>a)= f(x) / prob(x ž a)   Proof: definition of conditional prob= Joint / Marginal

If x µ N(. , 5²) and ! =(a– .)/5 =(standardized a, where a=the truncation point).
The density of **truncated normal** is f(x|x>a)= f(x) , 5 [1 • F(!)]
Note that the extra 5 in the denominator is from Jacobian of transf.
f(z)dz=f(x)dx must hold for any transformation from f(x) to f(z) to make sense.
hence f(x)=f(z) (dz/dx). Here z= (x– .)/5 means dz=dx/5. So (dz/dx)=(1/5)

THM 20.2 E(x| x>a)= . + 5 - , where if we keep RHS x>a, - =9Î [1–F]
where if we keep LHS x<a - =9Î [F]
- is called inverse-Mills-ratio and 9/[1–F] is also called the Hazard function for the dist'n.
Var(x| x>a)=5# [1-$] where $ is a shrinkage factor = - (- – !) Note 0<$<1 for all ! >0.
It is not surprising that the smaller range of the data reduces the variance.
Why hazard? death is certain, so any CDF - [0,1] can be used as a good for prob of death.
As time on horizontal axis Å we are all dead so prob Ä 1.
1–CDF (cumulative distribution function) is prob of survival=[1–F].
Length of life between t and t+ ? t is prob of death at t= hazard rate 9Î[1–F].
Examples of application of Thm:
(1) uniform dist'n has mean 0.5 and truncated uniform which keeps only x>(1/3)) has mean 0.66.
Thus the mean has increased if RHS of truncation point a=1/3 is Kept (LHS thrown out).
(2) Truncated lognormal: y=ln x has normal dist N(. , 5)
x=exp(y). E(x)=E(exp(y))=exp(. +0.5 5#)

Truncated regression model . = " "B3 (data on LHS and RHS of regression, both may be missing)
C3= " "B3 + %i where %i µ N(0,5#)
C3|B3 µ N(" "B3, 5#)
applying Thm 20.2 E(x| x>a)= . + 5 - here we have
E(C3| y3>a)= . + 5 - = " "B3 + 5 -
partial Š`E( | ) , `B3 ‹ = " +Š5 times (d - /d !)(d ! ÎdB3) ‹ by the chain rule.
The chain rule is needed because ! =(a–" "B3)Î5 does depend on B3.
Hence [d E(|)ÎdB3] = " (1 • $) where marginal effect is attenuated (reduced).
Similarly Var(C3| C3>a)=5#[1 • $] where $= - ²– ! - the shrinkage factor.
Olsen (1978) reparameterized to get rid of 5 in the denominators by defining
) =1/5 and #=" Î5. Then the log likelihood becomes more manageable.
Example of earnings eq. with marginal effect " (1 • $)

<u>**Censoring** is of the Dependent variable only:</u> This involves Winsorization of the data!
Tobin's study of auto demand uses censoring. There are many data points for which the dependent variable =0, is set (Winsorized at the limit value) even though it may be negative.
A "house full" performance means demand > seats available ¾ censored dep.LHS var.data missing.
There is Winsorization at the upper limit, values are simply replaced by upper limit values.
Other examples: durable goods purchase, extramarital affairs, arrests after release,
vacation expenditures, expenditures on commodity groups. (page 959)

THM 20.3: Moments of censored normal C‡ µ N(. ,5#)

Define new r.v. y from y^* as follows: y = a, if $y^* \le a$; and y = y^*, otherwise.
This is Winsorization (replacement of values in the **left tail** by the limiting value a). Then

$E(y) = \Phi a + (1 - \Phi)(\mu + \sigma\lambda)$

$var(y) = \sigma^2 (1 - \Phi)[(1 - \delta) + (\alpha - \lambda)^2 \Phi]$, where $\delta = \lambda^2 - \lambda\alpha$, $\lambda = \phi/(1 - \Phi)$

If the **right tail** is Winsorized ("house-full" events) then $E(y) = (1 - \Phi)a + \Phi(\mu + \sigma\lambda)$

The corresponding regression model is called **Tobit** model (in honor of Prof. Tobin)
Example 20.10 p 964 dep. var= hours worked by wives.
regressors are: prob of divorce, small kids, relative wage, etc.
Simple OLS regression gives reasonable estimates of sophisticated marginal effects.

<u>What can be WRONG with Tobit Model Specification Using Building-Fire Example:</u>
y_i=loss due to fire in a building, we are interested in the effect of age of the house.
If there is no fire, loss is zero. This is a REAL observation, like buying car where nothing (=0) was spent on a car.
$\partial prob(y_i>0)/\partial age > 0$ for old houses since old houses are more likely to have fires.
i.e. prob of positive loss from fire ↑ as age ↑
However, deriv. of expectation (distinguished from deriv. of prob) is Negative!
$\partial E(y_i | y_i>0)/\partial age < 0$ for old houses since old houses are more likely to have fires.
i.e. given that fire occurred, Expected loss ↓ as age ↑ older buildings are less expensive, and there is less to lose in a fire.
Note that positive coeff. in Tobit means both partials above MUST be OF the SAME sign.
The partial of the probability wrt x and partial of expected value wrt x need not be equal in examples like the house fires. Partial for prob of fire is positive and partial for expected loss is negative. To force the two to be equal is an unnecessary restriction of the Tobit (censored regression) model.
A more general model should have **Decision equation** $Pr(y^*>0) = \Phi(\gamma' z_i)$

$Pr(y^* \le 0) = 1 - \Phi(\gamma' z_i)$

then for nonlimit observations $E(y_i \mid z_i=1) = \beta' x_i + \sigma\lambda_i$

How to **test the restrictions** of the Tobit model?
(i) **Hetero**scedasticity may be handled directly.
(ii) For **misspecification** noted above Fin and Schmidt (1984 Rev.Eco.Stats, p.174) suggest a Lkhd ratio st'c is $-2[\ln L_T - (\ln L_P + \ln L_{TR})]$
subscript T for L is likelihood of Tobit, subscr P is for probit and sub TR for Trunc. regr.
one would have to estimate Truncated regr, probit and Tobit (ALL 3) to use this test.
Generalized Residual: p 883 of Greene

Binary choice model $\partial \ln L / \partial \beta = \sum_{i=1}^{n} (y_i - \Lambda_i) x_i = 0$ is the first order condition,

where Λ_i is Prob (Y=1) = $\exp(\beta' x)/[1 + \exp(\beta' x)]$, which is the CDF of logistic
here we can interpret $(y_i - \Lambda_i)$ as a generalized RESIDUAL.
Thus the partial of the log-Lkhd w.r.t. the intercept term (x_i=1 as the regressor) can be interpreted as a residual.
(iii) Testing **normality** assumption. (can be done by using 3rd and 4th moments)
Pagan and Vella (1989 J of App Etr supple) define conditional moment (CM) tests
$e_i = y_i - b' x_i$ where b= OLS estimator $(X'X)^{-1}(X'y)$
if variables are not erroneously omitted from regression $r_1 = (1/n)\sum z_i e_i$
$r_2 = (1/n)\sum z_i (e_i^2 - s^2)$ where $s^2 = e'e/n$ i.e. homoscedastic disturbances

<3=(1/n)D(vec)$_i$ where (vec) is a 2×1 vector with (e$_i^{\$}$ and e$_i^4$ –3 s$^{\%}$) i.e. errors are normally dist'ed.
the CM tests check these <$_i$=0.
Note that <$_3$ does not have the multiplier z$_3$

INCIDENTAL TRUNCATION **sample selection** model (Greene p. 974)
incidental truncation means non-random selection of data.
Assume that we are interested in incomes y (but in data collection we ignore low net worth people)
Wealth w is a related variable w=net worth. If the data are only for w>500,000 wealthy folks.
The study of incomes y here is not good, since the wealthy are likely to have a high y also.
f(y,w) is bivariate dist'n. It is truncated by including only w>a. The density f(y,w)
will have to be divided by the prob. that Pr(w>a) to get the correct truncated density here.

Where does inverse Mills ratio come from? It is in the mean and var. of incidentally truncated bivariate normal distributions? See below.

THM 20.4 E(y| z>a)=.c + 3 5$_y$- where -=9Î [1–F] if we keep RHS x>a
where -=9Î [F] if we keep LHS x<a
- is called inverse Mills ratio and 9/[1–F] . It is evaluated at (a–.$_z$)Î 5$_z$
It is also called the Hazard function for the dist'n.
Var(y| z>a)=5$^{\#}$[1–3^2 $] where $ is a shrinkage factor = -(--!) Note 0<$<1 for all !.
this is similar to theorem 20.2 except that we have the rho 3 present in both
Since |3|<1 by definition of correlation, the variance is shrunk (attenuated) even further!

Sample selection Á Simple Truncation
Note sample selection is different from simple truncation because
sample selection has one more variable (e.g. w for wealth or z from selection eq.) is involved.
Select only high net worth w>$500,000 people is a sample selection mechanism.

Example of female labor supply eq.
Two equations, Wage= f(age, education, children) and
Hours worked =f(wage, marital status, small children)
where desired working hours are observed only for working women.
b truncation of hours data, ¿ we have no data on hours for nonworking women.

selection variable z$_3^*$=# "w$_i$ + u$_i$, where z$_3^*$ is unobservable but z$_3$=1 if z$_3^*$>0 and z$_3$=0 otherwise.
probit eq. is Pr(z$_3$=1)= F(# "w$_i$)
Pr(z$_3$=0)= 1–F(# "w$_i$)
regression model: C$_3$= ""B$_3$ + %$_i$ is observed only if z$_3$=1
(u$_i$, %$_i$) µ Bivariate Normal(0, 0, 1, 5$_\%$, 3)
E(C$_3$ | z$_3$=1) = ""x + 3 5$_\%$ -(# "w)
Heckman 2-step procedure is (1) estimate probit by ML to estimate #
compute for each i density 9(# "w$_i$), F(# "w$_i$), and
-(# "w)=the inverse-Mills-ratio.
Estimate $ = S(S + w_i$) for variance term
($ is used in heterosc. adjustment)
(2) 2nd step is to regress C$_3$ on B$_3$ and -$_3$. Coeff of -$_3$ is 35$_\%$ (NOTE: coeff is not just 3 or 5$_\%$)
The errors of this regr. are heteros. with variance=5$_\%^2$(1–3^2 $$_3$)

Some extensions of selectivity models:
Treatment Effects: Self Selection problem. typical indivi who goes to college will earn more whether or no he goes to college! treatment=go to college
If so, the effect of (college) treatment is overestimated.

<u>What is the difference between **sample** selectioon and **self** selection?</u>
crit = criterion function tells us why data are missing.
If crit is written in "reduced form," we have sample selection.
If crit is written in **S**TRUCTURA**L** form, we have **S**E**L**F selection.
(aid to memory: it is selfish to own a big structure mansion as one's home!self sel=str eq.)
earnings = "'B$_3$ + $ C$_3$ + %$_i$ where C$_3$ is dummy variable =1 if went to college
this is a structural equation. Indivi. going to college will have high earnings,
even if the indivi did not go to college. ¾ This is self-selection to attend college.
the structural formulation. Greene p 982 has formualtion of this as a participation eq.
C_3^* = # 'w$_i$ + u$_i$ where C$_3$=1 if C_3^*>0 (ie person participates in the study, ie, goes to college)
It can be shown that simple OLS ignoring the selectivity-type correction underestiamtes.
here the additional regressor is –9$_3$ Î[1–F$_3$] for the non-college C$_3$=0 case.
the difference in earnings between participants and non-participants is
E(C$_3$|C$_3$=1) • E(C$_3$|C$_3$=0) œ $ + 3 5$_\%$ [9$_3$ / F$_3$Ð"–F$_3$)]
Problems: (1) normality assumption, robust approaches have been suggested.
Terza (1995, J of Etr.) has detailed derivation involving Poisson model. Greene p 982 summarizes.

§ 20.5 DURATION models
Typical data are cross section of durations of unemployment (of life despite a disease).
Censoring is pervasive is these studies.
Given that person has survived till time t, what is the prob that he will die by time=t+?t?
die(t,?)=Pr(t Ÿ age to death Ÿ t+?)
hazard rate -›= lim$_{?Ä!}$ dieÐt,?) / ? = lim$_{?Ä!}$ F(t+?) • F(t) / ?[1 • F(t)] œ lim$_{?Ä!}$ F(t+?) • F(t) / ?S(t) = f(t) / S(t)
where S(t) denotes the survival function. (Death is CDF and survival is 1–CDF)
Hazard function -(t)= • d ln S(t) / dt Note that f(t)=S(t)-(t) holds.
Also, integrated hazard (used later for assessment of the overall fit)
A(t)=' $^>_!$ -Ðs)ds, satisfies S(t)= exp(–A(t)), A(t)= • ln S(t),

Assume that -(t)=- or that it remains fixed over time (no memeory)
Then we can solve the differential equation -= • d ln S(t) / dt whose solution is
ln S(t)= constant • - t, ¾ S(t)= const of integration times exp(–-t).
Since S(0)=1, at time zero one is guaranteed to survive, cosnt of integration =1.
Finally the solution is S(t)=exp(–-t) This is exponential distribution whose mean is (1/-)
that is: E(t)= t̄ œ 1/-. thus to estimate - we simply use 1/ t̄
Some survival distributions are:
<u>Exponential</u> dist'n has hazard function -(t)=-, survival S(t)=exp(–-t).
Note we have e-to-the-power - times t not -(t) which may change with t.

<u>Weibull</u> dist. -(t)=-p (-t)$^{p-1}$ and S(t)=exp(–[-t]^p) additional parameter is p for Weibull.
If additional regressors are present, E(t|B$_3$)=exp("'B$_3$)>[(1/p)+1], note that p=1 for exponential
the shape of hazard function for exponential and Weibull is monotonic (downward)
see page 989 of Greene for similar expressions for log-normal and log-logistic.

the regression for log-normal and log-logistic is E(ln t|B3)="‶B3

S(t)= 1 , [1+(‑t)^7]

page 992 has a picture of hazard functions for these 4 distributions,
shape for lognormal and log-logistic is up-and-down

Testing: usual estimation is by ML so LkhdRatio, Lagr. Multiplier and Wald tests are available.
But for assessing overall fit, we need to consider integrated hazard function A
By "fundamental prob. transform," if z=CDF(t)=F(t), then t= F$^{–"}$(z)
Now let the CDF of z be F(z) and substitue z=F(t)=F(F$^{–"}$(z)) to yield F(z)=FF$^{–"}$z=z
This CDF for z between 0 to 1 is such that its CDF equals the value of z between 0 to 1.
This CDF will lie on the 45 degree line for z ‑ [0,1]. Hence the distribution must be
Uniform distribution. Do the sample survival rates look like a sample from a Uniform?
Instead, one defines the generalized residual as ‰ œ A(t|)) œ • ln S(t|)) ¾ exp(–‰)=S(t|))
CDF (‰)=F(‰)=1 • exp(–‰) which is the cdf of an exponential variable with parameter 1
(unit paramter exponential). Its moments are well known to be E(‰$^<$Ñ=r!
Hence one can test if the mean of residuals is 1 and the other central moments are 1,2 and 9

Testing for Weibull also can use conditional moment tests E(‰^{2})=2, E(‰^{3})=6
E(‰^{4})=24 etc. See page 004-995 for details of Pagan-Vella's tests

Heterogeneity problem:
Individuals usually do have distinct characteristics which cannot be completely modeled.
Kaplan-Meier nonparametric approach is immune to this, where an empirical estimate of
survival function is directly obtained from data as § (T_k) œ $\frac{n_5–2_5}{8_5}$ A2/</ n_k=observed duration is at
least T_k and h_k=number of deaths by time T_k
§ = (n_k • h_k)Î n_k needs adjustments for censored observations.
There is left censoring, right censoring and random censoring
A better description of Kaplan-Meier is in V&R book on S language

Estimating survivor curve for uncensored data is easy. Just take 1–empirical CDF.
For conf. intervals use Binomial. Variance is §(1–§)În œ r(n–r)În$^\$$ß A2/</ r=those at risk.
d_3=no. of deaths. Consider a set of intervals over [0, _) deined by I_3=[t_3, $t_{3€"}$)
:$_3$=prob of surviving the interval I_3
prob of surviving until time t_3 is S(t_i)=$\prod_{4œ!}^{3–"}$:$_4$ ¸ $\prod_{4œ!}^{3–"}$ $\frac{<\text{Ð}t_3) • d_3}{r(t_3)}$ the product may be defined
only for non-unity cases (No need to multiply by 1). then it is Kaplan-Meier estimator.
A command in S can look like this:

```
temp<-survfit(Surv(time,cens)~treat, conf.type="log-log")
```

Cox-Proportiional Hazard Model page 279 of Venables & Ripley
Cox(1972) proposed less parametric approach
h(t)= h$_!$(t) exp("‶x), where h$_!$(t) is called the baseline hazard (we are not intereseted in it)
i-th person may live or die at time t_4
Assume that one person dies at time t_j. Then conditional on this event, the prob that i-th person
died is h(t) , ’D› Indicator(T_i t)h(t)“. Note that in this formula the baseline hazard cancels.
We need the censoring to be uninformative. Also there is a problem of ties.