204B Problem Set 4

Clinton Tepper

02/01/2017

914-589-5370

clinton.tepper@gmail.com

Problem 1

a

- We start with the inner integral

$$
\begin{aligned}
I =& \int_t^\infty \left(u - E\left[s_i\right]\right) g\left(u\right) du \\
I =& \int_t^\infty u g\left(u\right) du - \int_t^\infty E\left[s_i\right] g\left(u\right) du \\
I =& E\left[s_i,\ s_i \geq t\right] - E\left[s_i\right] p\left(s_i \geq t\right) \\
I =& E\left[s_i | s_i \geq t\right] p\left(s_i \geq t\right) - E\left[s_i\right] p\left(s_i \geq t\right) \\
I =& E\left[s_i | s_i \geq t\right] p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) + E\left[s_i | s_i \geq t\right] p\left(s_i \geq t\right) p\left(s_i \geq t\right) - E\left[s_i\right] p\left(s_i \geq t\right) \\
I =& E\left[s_i | s_i \geq t\right] p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) + E\left[s_i | s_i \geq t\right] p\left(s_i \geq t\right) p\left(s_i \geq t\right) \\
& - E\left[s_i | s_i \geq t\right] p\left(s_i \geq t\right) p\left(s_i \geq t\right) - E\left[s_i | s_i \geq t\right]\left(1 - p\left(s_i \geq t\right)\right) p\left(s_i \geq t\right) \\
I =& E\left[s_i | s_i \geq t\right] p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) - E\left[s_i | s_i < t\right]\left(1 - p\left(s_i \geq t\right)\right) p\left(s_i \geq t\right) \\
I =& \left(E\left[s_i | s_i \geq t\right] - E\left[s_i | s_i < t\right]\right) p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) \checkmark
\end{aligned}
$$

b

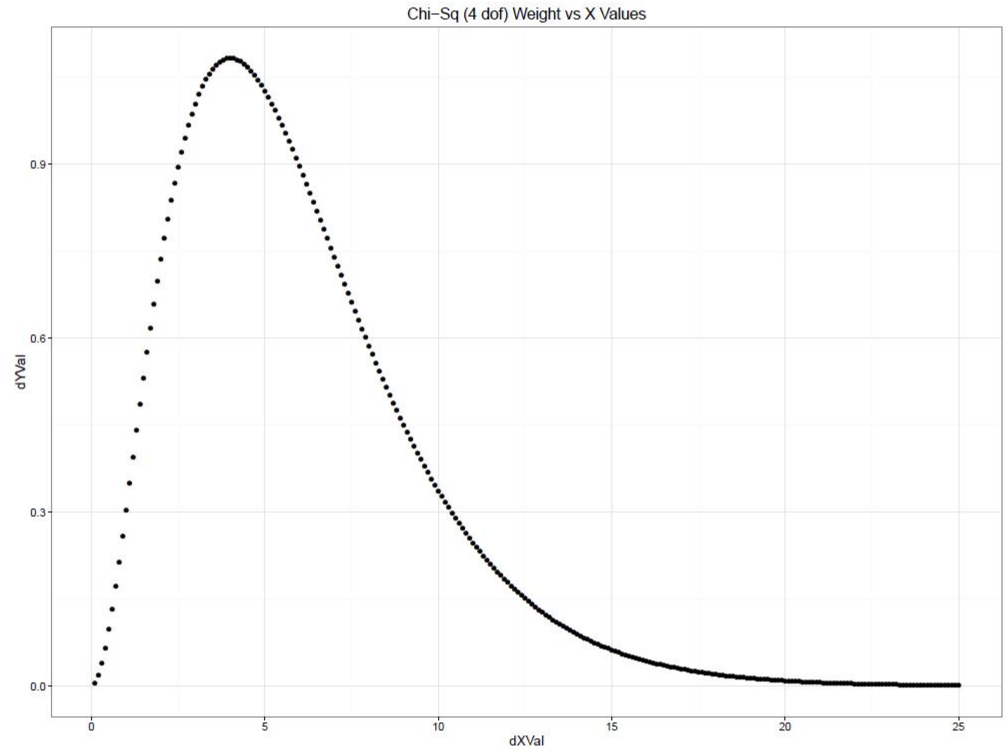
- The general formula for the mean of a truncated normal density is $E\left(X | a < X < b\right) = \mu + \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma$.
- Therefore, we can write the expectations as:

$$
\begin{aligned}
I =& \left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \frac{-\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)}\right) \sigma p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) \\
I =& \left(\frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{t-\mu}{\sigma}\right)} - \frac{-\phi\left(\frac{t-\mu}{\sigma}\right)}{\Phi\left(\frac{t-\mu}{\sigma}\right)}\right) \sigma p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) \\
I =& \left(\frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{p\left(s_i \geq t\right)} - \frac{-\phi\left(\frac{t-\mu}{\sigma}\right)}{1 - p\left(s_i \geq t\right)}\right) \sigma p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) \\
I =& \frac{\phi\left(\frac{t-\mu}{\sigma}\right)\left(1 - p\left(s_i \geq t\right)\right) - \phi\left(\frac{t-\mu}{\sigma}\right) p\left(s_i \geq t\right)}{p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right)} \sigma p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right) \\
I =& \left[\phi\left(\frac{t-\mu}{\sigma}\right)\left(1 - p\left(s_i \geq t\right)\right) - \phi\left(\frac{t-\mu}{\sigma}\right) p\left(s_i \geq t\right)\right] \sigma \\
I =& \phi\left(\frac{t-\mu}{\sigma}\right) \sigma \checkmark
\end{aligned}
$$

c

- The weights are naturally given by $u_t = \left(E\left[s_i|s_i \geq t\right] - E\left[s_i|s_i < t\right]\right) p\left(s_i \geq t\right)\left(1 - p\left(s_i \geq t\right)\right)$.

```r
require(devtools);
require(DataAnalytics);
require(plyr);
require(ggplot2);

set.seed(11);
#allows for reproducability

#A function to get the truncated expectation given a distribution and break points
#In: The distribution and break points
#Out: The expectation
TruncMean = function(Dist, dLo, dHi) {
    if (missing(dLo)) dLo = -Inf;
    if (missing(dHi)) dHi = Inf;
    return(integrate(function(x) x * Dist(x), lower = dLo, upper = dHi)$val);
}

#A function to get the weights using the weighted derivitive interpretation of a regression
#In: The distribution, cumulative distribution, and point for the weight
#Out: The expectation
regWeight = function(Dist, CumDist, dT, dLowerBound, dUpperBound) {
    if (missing(dLowerBound)) dLowerBound = -Inf;
    if (missing(dUpperBound)) dUpperBound = Inf;
    dProb = 1-CumDist(dT);
    return((TruncMean(Dist, dLo = dT, dHi = dUpperBound)/dProb -
        TruncMean(Dist, dLo = dLowerBound, dHi = dT)/(1-dProb)) * dProb * (1 - dProb));
}

#########################################Script Entry Point###########################
Dist = function(x) dchisq(x, 4);
CumDist = function(x) pchisq(x, 4);
dLo = 0.1;
dHi = 25;
dInterval = .1;

dXVal = seq(from = dLo, to = dHi, by = dInterval);
dYVal = sapply(dXVal, function(x) regWeight(Dist = Dist, CumDist = CumDist, dT = x, dLowerBound = 0));

ggplot(data.frame(dXVal, dYVal), aes(x = dXVal, y = dYVal)) +
    geom_point() + theme_bw() +
    ggtitle("Chi-Sq (4 dof) Weight vs X Values");
```



- We conclude by noting how the graph indicates that higher weight is provided to values closer to the mean.

Problem 2

a

- Coefficients to alternative models, like the Probit model, do not correspond linearly to the magnitude of the effect.
  - "But the Probit Coefficients...do not give us the size of the effect...until we feed them back into the normal CDF (though they do have the right sign)." (pg. 98)
- With regards to conditioning on positive effects, a censoring model such as the Tobit model is hard to interpret when the variables are not actually censored.
  - "Healthcare expenditure is really zero for some people; this is not a statistical artifact or due to some kind of censoring. So the notion of ... a potentially negative Y_i^* is hard to grasp." (pg. 100)
- The relationship between the estimated parameters in models like the Tobit model and the causal effect of interest can depend on the distributional assumptions of the model.

– "A second problem is that the link between the parameter β_1^* in the latent model and causal assumptions of the observed outcome, Y_i, turn on distributional assumptions about [a] latent variable." (pg. 100)

b

- The Probit model is given by $\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)$
- The likelihood is given by $L = \prod_I \left(Y_i\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) + (1-Y_i)\left(1-\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)\right)\right)$
- Noting that there is a single independent variable, the log-likelihood is given by: $l = \sum_I log\left(Y_i\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) + (1-Y_i)\left(1-\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)\right)\right)$
- Let η_Y equal $\sum_I Y_i$, η_X equal $\sum_I X_i$, η_{XY} equal $\sum_I X_iY_i$, and N be the total number of samples.
  - Then the MLE for β_1 is characterized as:

$$FOC: 0 = \sum_I \frac{\left(\frac{X_i}{\sigma_v}\right)\left[Y_i\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) - (1-Y_i)\,\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)\right]}{Y_i\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) + (1-Y_i)\left(1-\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)\right)}$$

$$\begin{aligned}
0 =& \eta_{XY}\frac{\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)}{\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)} - (\eta_X - \eta_{XY})\frac{\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)}{\left(1-\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\right)}\\
0 =& \eta_{XY}\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\left(1-\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\right) - (\eta_X-\eta_{XY})\,\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\\
0 =& \eta_{XY}\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right) - \eta_{XY}\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\\
&- n_X\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right) + n_{XY}\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\\
0 =& \eta_{XY}\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right) - \eta_X\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\\
0 =& \eta_{XY} - \eta_X\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)\\
\beta_1 =& \sigma_\nu\Phi^{-1}\left(\frac{\eta_{XY}}{\eta_X}\right) - \beta_0
\end{aligned}$$

  - If $\beta = 0$,this reduces to $\beta_1 = \Phi^{-1}\left(\frac{\eta_{XY}}{\eta_X}\right)$
  - Note that $\frac{\eta_{XY}}{\eta_X} = p\,(Y|X)$, so $\beta_1 = \Phi^{-1}\left(p\,(Y|X)\right) - \beta_0$

- The MLE for β_0 is characterized by:

$$
\begin{aligned}
FOC: 0 = & \sum_I \frac{Y_i \phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) - (1-Y_i)\,\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)}{Y_i\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) + (1-Y_i)\left(1-\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)\right)} \\
0 = & \sum_I \left[\frac{Y_i\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)}{\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)} - \frac{(1-Y_i)\,\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)}{1-\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)}\right] \\
0 = & \frac{n_{XY}\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)}{\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)} + \frac{(\eta_Y-\eta_{XY})\,\phi\left(\frac{\beta_0}{\sigma_\nu}\right)}{\Phi\left(\frac{\beta_0}{\sigma_\nu}\right)} \\
& - \frac{(\eta_X-\eta_{XY})\,\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)}{1-\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)} - \frac{(N-\eta_X-(\eta_Y-\eta_{XY}))\,\phi\left(\frac{\beta_0}{\sigma_\nu}\right)}{1-\Phi\left(\frac{\beta_0}{\sigma_\nu}\right)}
\end{aligned}
$$

  - Note that in this case, the equation does not reduce cleanly as there is no indicator for X outside of the derivitive.

- The FOC for the MLE for σ_v is characterized as follows:

$$
\begin{aligned}
0 = & \sum_I \frac{\left(-\frac{X_i\beta_1+\beta_0}{\sigma_v^2}\right)\left[Y_i\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) - (1-Y_i)\,\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)\right]}{Y_i\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right) + (1-Y_i)\left(1-\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)\right)} \\
0 = & \sum_I [X_i\beta_1+\beta_0]\left[\frac{Y_i\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)}{\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)} - \frac{(1-Y_i)\,\phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)}{1-\Phi\left(\frac{X_i\beta_1+\beta_0}{\sigma_v}\right)}\right] \\
0 = & \frac{(\beta_1+\beta_0)\,n_{XY}\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)}{\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)} + \frac{\beta_0\,(\eta_Y-\eta_{XY})\,\phi\left(\frac{\beta_0}{\sigma_\nu}\right)}{\Phi\left(\frac{\beta_0}{\sigma_\nu}\right)} \\
& - \frac{(\beta_1+\beta_0)\,(\eta_X-\eta_{XY})\,\phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)}{1-\Phi\left(\frac{\beta_1+\beta_0}{\sigma_v}\right)} - \frac{\beta_0\,(N-\eta_X-(\eta_Y-\eta_{XY}))\,\phi\left(\frac{\beta_0}{\sigma_\nu}\right)}{1-\Phi\left(\frac{\beta_0}{\sigma_\nu}\right)}
\end{aligned}
$$

  - Again the equation does not reduce cleanly. Also note that there is an identification issue, as the FOC for σ_v can be written as a linear combination of the FOC for β_1 and β_0. This may be why the model frequently is written without σ_v, the implict substitution being $\beta_1' = \frac{\beta_1}{\sigma_v}$ and $\beta_0' = \frac{\beta_0}{\sigma_v}$.

- In a linear regression, we have $\beta = (X'X)^{-1} X'Y$. Then $X'X = \begin{bmatrix} N & \eta_X \\ \eta_X & \eta_X \end{bmatrix}$ and $X'Y = \begin{bmatrix} \eta_Y \\ \eta_{XY} \end{bmatrix}$.

- Therefore, $\beta = \begin{bmatrix} N & \eta_X \\ \eta_X & \eta_X \end{bmatrix}^{-1} \begin{bmatrix} \eta_Y \\ \eta_{XY} \end{bmatrix}$.
- If we assume $\beta_0 = 0$, then the above equation reduces to $\beta_1 = \frac{\eta_{XY}}{\eta_X}$
- In either case, $\beta_{OLS} \neq \beta_{Probit}$

c

- With continuous effects, the model remains $\Phi\left(X_i\beta_1 + \beta_0\right)$.
- Presumably the coefficients are estimated using the traditional MLE method and solving

$$\beta = argmax_\beta \left\{ \sum_I log\left(Y_i\Phi\left(X_i\beta_1 + \beta_0\right) + \left(1 - Y_i\right)\left(1 - \Phi\left(X_i\beta_1 + \beta_0\right)\right)\right) \right\}$$

- The marginal impact at any given x is given by $\phi\left(X_i\beta_1 + \beta_0\right)\beta_1$
- The average marginal impact for X could be calculated parametrically from the expectation, given by $\int_{-\infty}^{\infty} x\phi\left(x\beta_1 + \beta_0\right)\beta_1 dx = \beta_1 E\left[\phi\left(x\beta_1 + \beta_0\right)\right] = \beta_1\beta_0$.

Problem 3

- In the linear homoskedastic model, we can write $V\left(Y|X\right) = E\left[\left(Y - X\beta\right)'\left(Y - X\beta\right)\right] = \sigma_\varepsilon^2$. We can also re-write $V\left(Y\right) = V\left(E(Y|X)\right) + E\left(V\left(Y|X\right)\right) = \beta'\sigma_x'\sigma_x\beta + \sigma_\varepsilon^2$
- We can now get a more intuitive understanding of ρ^2 by writing out its estimator, $\rho^2 = 1 - \frac{SSE}{(n-p_\lambda)V(Y)} = 1 - \frac{n-1}{n-p_\gamma}\frac{SSE}{SST}$. Here we assume $p_\lambda = k$, the number of columns in the X matrix. It is easy to see how the statistic is model independent. Moreover, if we define a correspondence $G\left(X\right) : \Re^k \mapsto \Re^{k'}$ which transforms X, then we have $\rho^2 = 1 - \frac{n-1}{n-p_\gamma}\frac{SSE}{SST} = 1 - \frac{n-1}{n-p'_\gamma}\frac{(Y-G(X)\beta)'(Y-G(X)\beta)}{Y'Y} = 1 - \frac{n-1}{n-p'_\gamma}\frac{(Y-G(X)\beta)'(Y-G(X)\beta)}{Y'Y}$. Thus outside of the p_λ adjustment, we can expect the measurement to function in a model independent manner.
- Note that the **estimate** for ρ^2 is unbounded from below, as $\frac{n-1}{n-p_\gamma}$ is unbounded from above. Therefore, the **estimate** of ρ^2 ranges from $(-\infty, 1]$, even though the **true value** of $\rho^2 = 1 - \frac{V(Y|X)}{V(Y)}$ ranges from $[0, 1]$.
- Note that R^2 as conventionally defined, $R^2 = 1 - \frac{SSE}{SST}$, will function as a biased estimator of ρ^2 unless $p_\lambda = 1$, at which point the regression is just an unconditional mean of Y.

- The R^2 estimator is consistent, as shown by:

$$
\begin{aligned}
R^2 =& 1 - \frac{SSE}{SST} \\
R^2 =& 1 - \frac{(Y - X\beta)'(Y - X\beta)}{(Y - Y)'(Y - Y)} \\
R^2 =& 1 - \frac{n}{n}\frac{(Y - X\beta)'(Y - X\beta)}{(Y - Y)'(Y - Y)} \\
R^2 =& 1 - \frac{\sigma_\varepsilon^2}{V(Y)} \text{ (By WLLN)}
\end{aligned}
$$