Final Exam Question 6

Consider the Poisson Regression Model:

$$Y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

and recall that $E[Y_i] = \lambda_i$ and $Var(Y_i) = \lambda_i$.
As discussed in class, one of the concerns with the Poisson regression model is the requirement that the mean and variance be equal, and this can cause problems if the data is over-dispersed.
One solution is to replace the Poisson distribution with a version of the Negative Binomial distribution

$$P(Y_i = y) = \frac{\Gamma(y+r)}{y!\Gamma(r)} p_i^r (1-p_i)^y$$

which is parametrized by p_i and r, with $0 \leq p_i \leq 1$ and $r > 0$ (traditionally when you first learn the Negative Binomial distribution you learn that r has to be an integer, but in reality it can be any positive real number).
In this parameterization, $E[Y_i] = \frac{r(1-p_i)}{p_i}$ and $Var(Y_i) = \frac{r(1-p_i)}{p_i^2}$

(a) If we plan to replace the Poisson distribution with the above Negative Binomial distribution in our regression, we would like to keep the same link function, namely that $\log(\lambda_i) = \log(E[Y_i]) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$.

Equating the mean of the Poisson with the mean of the Negative Binomial, express p_i as a function of λ_i and r.

(b) Using your answer to part (a), rewrite the $Var(Y_i)$ in the Negative Binomial case in terms of λ_i and r.

(c) The model introduced in Question 4 contains two Poisson regressions, one for the Home Goals, and one for the Away Goals. Replace each of these with Negative Binomial's and use your answer to part(a) to connect the parameters of the Negative Binomial to the existing link functions for $\log(\lambda_{ij})$ and $\log(\theta_{ij})$.

(d) Run the Negative Binomial model for the EPL data, and answer the same questions as posed in Question 4 (b) through (f) and Question 5 (h) through (l). Are your answers any different?

Show all working.