1. (a) We have \( E[\tilde{\beta}] = E[\beta + (Z'AX)^{-1}Z'\mathbf{u}] = \beta + (Z'AX)^{-1}Z'\text{AE}[\mathbf{u}] = \beta \)
using \( Z, A \) and nonstochastic and \( E[\mathbf{u}] = 0 \).
And \( V[\tilde{\beta}] = E[(\tilde{\beta} - E[\tilde{\beta}]) (\tilde{\beta} - E[\tilde{\beta}])'] = E[(Z'AX)^{-1}Z'\mathbf{u} ((Z'AX)^{-1}Z'\mathbf{u})'] \)
\( = (Z'AX)^{-1}Z'A\Sigma AZ(X'AZ)^{-1}. \)

(b) We have \( \hat{\beta} = (Z'AX)^{-1}Z'A(X\beta + \mathbf{u}) = \beta + (Z'AX)^{-1}Z'\mathbf{u}. \) So
\[
\hat{\beta} = \beta + (N^{-1}Z'AX)^{-1}N^{-1}Z'A \mathbf{u}
\]
\[
\xrightarrow{p} \beta + (\text{plim} N^{-1}Z'AX)^{-1} \text{plim} N^{-1}Z'\mathbf{u}
\]
assuming first plim is finite non-zero and second is zero (which essentially requires \( E[u|Z] = 0 \)).

(c) We have
\[
\sqrt{N}(\hat{\beta} - \beta) = (N^{-1}Z'AX)^{-1} \frac{1}{\sqrt{N}} Z'\mathbf{u}
\]
\[
\xrightarrow{d} (\text{plim} N^{-1}Z'AX)^{-1} \times \mathcal{N}[0, \mathbf{B}]
\]
\[
\xrightarrow{p} \mathcal{N}[0, (\text{plim} N^{-1}Z'AX)^{-1} \mathbf{B} (\text{plim} N^{-1}X'AZ)^{-1}]
\]
where it is assumed that relevant LLN and CLT can be applied, where
\[
\mathbf{B} = \lim V \left[ \frac{1}{\sqrt{N}} Z'\mathbf{u} \right] = \lim E \left[ N^{-1}Z'\mathbf{uu}'AZ \right] = \lim N^{-1}Z'\mathbf{A} \Sigma AZ.
\]

(d) Given heteroskedastic errors the White approach can be applied so
\[
\text{plim} N^{-1}Z'ADiag[\hat{\sigma}_i^2]AZ = \lim N^{-1}Z'\mathbf{A} \Sigma AZ, \text{ where } \hat{\sigma}_i = y_i - x'_i \hat{\beta},
\]
and
\[
\hat{\mathbf{V}}[\tilde{\beta}] = (Z'AX)^{-1}Z'ADiag[\hat{\sigma}_i^2]AZ(X'AZ)^{-1}.
\]

(e) Given \( A \) is diagonal symmetric \( A^{1/2} = \text{Diag}[\sqrt{\sigma_i}] \) satisfies \( A^{1/2}A^{1/2} = A \), so
\[
\hat{\beta} = \left[ X'AX \right]^{-1}X'\mathbf{y} = \left[ (X'\mathbf{A}^{1/2})(\mathbf{A}^{1/2}X) \right]^{-1}(X'\mathbf{A}^{1/2})(\mathbf{A}^{1/2}\mathbf{y})
\]
\[
= \left[ (\mathbf{A}^{1/2}X)'(\mathbf{A}^{1/2}X) \right]^{-1}(\mathbf{A}^{1/2}X)'(\mathbf{A}^{1/2}\mathbf{y})
\]
which is OLS of \( A^{1/2}\mathbf{y} \) on \( A^{1/2}X \), or OLS of \( \sqrt{\sigma_i}y_i \) on \( \sqrt{\sigma_i}x_i \).
2. (a) Here

\[\ln f(y) = 2\ln \lambda + \ln y - \lambda y \text{ and } \lambda = 2 \exp(-x'\beta) \text{ and } \ln \lambda = \ln 2 - x'\beta\]

\[\Rightarrow \ln L(\beta) = \sum_i \ln f(y_i) = \sum_i \{2 \ln 2 - 2x'_i\beta + \ln y_i - 2 \exp(-x'_i\beta)y_i\}.

(b) Here

\[\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_i (-2x_i + 2y_i \exp(-x'_i\beta)x_i)
= \sum_i 2 \times \{y_i \exp(-x'_i\beta) - 1\}x_i \text{ rearranging}
= \sum_i 2 \times \frac{y_i - \exp(x'_i\beta)}{\exp(x'_i\beta)}x_i \text{ multiplying by } \frac{\exp(x'_i\beta)}{\exp(x'_i\beta)}
\]

(c) Essential condition is \(E\left[y_i | x_i\right] = \exp(x'_i\beta)\) as then \(E\left[\frac{y_i - \exp(x'_i\beta)}{\exp(x'_i\beta)}x_i\right] = 0\) at \(\beta = \beta_0\) so then

\[E\left[\frac{\partial \ln L(\beta)}{\partial \beta} \bigg|_{\beta_0}\right] = 0.
\]

(d) For the MLE

\[\sqrt{N}(\hat{\beta} - \beta_0) \overset{d}{\rightarrow} N[0, A_0^{-1}]\]

where \(A_0 = -\lim E\left[\frac{1}{N} \frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta'} \bigg|_{\beta_0}\right]\) where

\[\frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta'} = \sum_i -2y_i \exp(-x'_i\beta)x_ix'_i
\]

\[E\left[\frac{\partial^2 \ln L(\beta)}{\partial \beta \partial \beta'} \bigg|_{\beta_0}\right] = \sum_i -2 \exp(x'_i\beta_0) \exp(-x'_i\beta_0)x_ix'_i = -\sum_i 2x_ix'_i
\]

so

\[\sqrt{N}(\hat{\beta} - \beta_0) \overset{d}{\rightarrow} N\left[0, \left(\lim 2N^{-1} \sum_i x_ix'_i\right)^{-1}\right].\]

(e) Use Newton-Raphson

\[\left(\hat{\beta}_{s+1} - \hat{\beta}_s\right) = -H_s^{-1}g_s = \left(\sum_i 2y_i \exp(-x'_i\hat{\beta}_s)x_ix'_i\right)^{-1}\sum_i 2\frac{y_i - \exp(x'_i\hat{\beta}_s)}{\exp(x'_i\hat{\beta}_s)}x_i
\]

and can cancel the 2’s. Or can take expected value of the hessian and use

\[\left(\hat{\beta}_{s+1} - \hat{\beta}_s\right) = -\left(E[H]_{\beta_s}\right)^{-1}g_s = \left(\sum_i 2y_i \exp(-x'_i\hat{\beta}_s)x_ix'_i\right)^{-1}\sum_i 2\frac{y_i - \exp(x'_i\hat{\beta}_s)}{\exp(x'_i\hat{\beta}_s)}x_i.
\]
3. (a) Here $u_{it}$ is causing problems, so do instrumental variables of $y_{it}$ on $x_{it}$ with instruments $z_{it}$. Here $z_{it}$ should be uncorrelated with $u_{it}$ (and $w_i$) and correlated with $x_{it}$.

(b) Actually same solution as part (a). Or can first difference to get rid of $w_i$ and then do IV.

(c) Estimation is based on $E[z_i(y_i - \Lambda(x_i'\beta))] = 0$.
If just-identified the estimator solves $\sum_i z_i(y_i - \Lambda(x_i'\beta)) = 0$.
If over-identified GMM estimator minimizes $[\sum_i z_i(y_i - \Lambda(x_i'\beta))]' W [\sum_i z_i(y_i - \Lambda(x_i'\beta))]$.

(d) Density for $i^{th}$ observation is $F_1(x_i'\beta_1)^{y_{1i}} \times F_2(x_i'\beta_2)^{y_{2i}} \times F_3(x_i'\beta_3)^{y_{3i}}$.
Log-likelihood is $\sum_i y_{1i} \ln F_1(x_i'\beta_1) + y_{2i} \ln F_2(x_i'\beta_2) + y_{3i} \ln F_3(x_i'\beta_3)$.

(e) For nonlinear estimators the optimal estimator has smallest asymptotic variance matrix among all consistent estimators.

4. (a) For logit, sign of coefficient gives sign of the marginal effect.
Expect fairpoor down with increase in income; up with increase in age; up with increase in chronic.
So the signs of the coefficients are as expected.

(b) Yes. Individually at level 0.05 as $p < 0.05$ using either one-sided or two-sided test. And jointly as overall Wald test ($\chi^2(2)$) has $p = 0.00$.

(c) A $1,000 increase in income is a one-unit increase in income.
Using $\partial \ln \Pr[y_i = 1|x_i] / \partial x = \Lambda(x_i'\beta)(1 - \Lambda(x_i'\beta))\beta$ evaluated at $\Lambda(x_i'\beta) = \bar{y}$ we have that probability of fairpoor falls by $.097 \times .903 \times .0236 = 0.0021$.
Or use rule of thumb. Falls by $.025 \times .0236 = 0.0059$.
Or log-odds ratio falls by $-.0236$. Or odds-ratio is $e^{-0.0236} = 0.978$ times that without income change.

(d) Consistency requires that $\Pr[y_i = 1|x_i] = \exp(x_i'\beta)/[1 + \exp(x_i'\beta)]$.
Efficiency requires correct density. Here density for each observations is necessarily Bernoulli, but we do additionally require independence over $i$ for correct joint density of all observations.

(e) Difference is in the standard errors. The robust se’s here are actually quite different for income, increasing from 0.0029 to 0.0037, a more than 30% increase.

(f) The model fit of logit is better than probit as the log-likelihood is $-1249.3 - -1253.9 = 4.6$ higher for logit. [This is a large difference actually, as a likelihood ratio test with one degree of freedom favors the more general model at level 0.05 if critical value is 3.84 or likelihood ratio change exceeds 3.84/2 = 1.92.]
Also using the pseudo-$R^2$ logit is favored as has higher pseudo-$R^2$. 