Brief Solutions

1. (a) Premultiplying both sides by $\begin{pmatrix} X_1' M_{x_2} X_1 & 0 \\ x_2' X_1 & x_2' x_2 \end{pmatrix}$ gives $X_1' M_{x_2} X_1 \hat{\beta}_1 = X_1' M_{x_2} y$ and $x_2' X_1 \hat{\beta}_1 + x_2' x_2 \hat{\beta}_2 = x_2' y$ and hence $\hat{\beta}_1 = (X_1' M_{x_2} X_1)^{-1} X_1' M_{x_2} y$, which is from the Frisch-Waught Theorem, and $\hat{\beta}_2 = \frac{x_2' y - x_2' X_1 (X_1' M_{x_2} X_1)^{-1} X_1' M_{x_2} y}{x_2' x_2}$, which is obtained by regressing $y - X_1 \hat{\beta}_1$ on x_2.

   (b) $\tilde{\beta}_1 = (X_1' X_1)^{-1} X_1' y = (X_1' X_1)^{-1} X_1' (X_1 \beta_1 + x_2 \beta_2 + \epsilon) = \beta_1 + (X_1' X_1)^{-1} X_1' x_2 \beta_2 + (X_1' X_1)^{-1} X_1' \epsilon$. In other words, $\tilde{\beta}_1$ is also consistent and is asymptotically equal to $\hat{\beta}_1$ if either $\beta_2 = 0$ or the canonical correlation between x_2 and the columns of X_1 is 0.

   (c) $\hat{\beta}_1 - \tilde{\beta}_1 = (X_1' M_{x_2} X_1)^{-1} [X_1' M_{x_2} - X_1' M_{x_2} X_1 (X_1' X_1)^{-1} X_1'] y = (X_1' M_{x_2} X_1)^{-1} X_1' M_{x_2} [I - X_1 (X_1' X_1)^{-1} X_1'] y = (X_1' M_{x_2} X_1)^{-1} X_1' M_{x_2} M_{X_1} y$. Under H_0, $\hat{\beta}_1 - \tilde{\beta}_1$ is asymptotically equivalent to $(X_1' M_{x_2} X_1)^{-1} X_1' M_{x_2} M_{X_1} \epsilon$, hence one can construct the test statistic $\frac{1}{\sigma^2} (\hat{\beta}_1 - \tilde{\beta}_1)' [X_1' M_{x_2} X_1 (X_1' M_{x_2} M_{X_1} M_{x_2} X_1)^{-1} X_1' M_{x_2} X_1] (\hat{\beta}_1 - \tilde{\beta}_1)$. The student is not required to point out what distribution the test statistic follows ($\chi^2(K)$ under some more conditions not given in the question).

2. See Page 3 in Lecture 6 for the first question. See Eq.(22) and Theorem 4.3 in Lecture 6.

3. (a) $plim \hat{b}_1 = \frac{Cov(x_{1i}, y_i)}{Var(x_{1i})} = 0.6$ and $plim \hat{b}_0 = E(y_i) - E(x_{1i}) b_1 = 0$.

   (b) The two R^2 are the same. Both are equal to the sample correlation between y_i and x_{1i}.

   (c) $plim (\hat{d}_1, \hat{d}_2)' = Var((x_{1i}, x_{2i})')^{-1} Cov((x_{1i}, x_{2i})', y_i) = (\frac{3}{7}, \frac{3}{7})'$, $plim R^2 = \frac{1}{Var(y_i)} Cov(y_i, (x_{1i}, x_{2i})) Var((x_{1i}, x_{2i})')^{-1} Cov((x_{1i}, x_{2i})', y_i) = \frac{6}{7}$.

   (d) There will be perfect collinearity. y_i can be written as a deterministic linear function of x_{1i}, x_{2i} and x_{3i}.

4. The limitations of linear probability model (LPM) are

   - It is possible to have probability point forecast, $x_i' \hat{\beta}$, outside the range of $[0, 1]$, which does not make sense for $P(y_i = 1 | X)$.

   - Point forecast works well only for observations close to the sample average of x_i.

- The probability of success is linearly related to the independent variables for all their possible values.

For logit or probit model, they relate the probability response to $x_i'\beta$ through a nonlinear function, i.e. $P(y_i = 1|X) = G(x_i'\beta)$, where $G(\cdot)$ is the cumulative distribution function (logistic for logit and standard normal for probit). Since cumulative distribution function is always between 0 and 1, probability point forecast will lie in sensible range. Moreover, the marginal effect of x_i on the probability response will depend on not only β, but also on $g(x_i'\beta)$, where $g(\cdot)$ is the probability density function. Since both the logistic and standard normal distribution have bell shape densities, the magnitude increase of x_i (i.e. $|x_i|$) will have diminishing marginal effect on the probability response. We can estimate logit or probit model by maximum likelihood. The slope coefficients will not have the same interpretation. For LPM, the slope can be interpreted as the marginal effect of x_i on the probability response. But for logit and probit model, the marginal effect also depend on the density function evaluated at $x_i'\beta$. The density functions are different for the two models.

5. (a) If IQ affects earning, not including IQ will lead to omitted variable bias. Since a person's level of education could be correlated with the person's IQ, the bias in the estimate of β_1 will not disappear as the sample size grows.

   (b) We can calculate the F statistic based on the SSRs from the restricted and unrestricted model, which is equal to $\frac{(2474.967-2426.078)/2}{2426.078/(852-9)} \approx 8.494$. Since the 5% critical value for $F_{2,843}$ is around 3, we know that *brthord* and *sibs* are jointly very significant. If there are more children, families may not be able to afford their education. For birth order, it could be that older children are given priority for higher education, and families may hit budget constraints and may not be able to afford as much education for children born later.

   (c) The 10% critical value for a student t distribution with 843 degrees of freedom is about -1.647. We can say that $\hat{v}$ is significantly different from zero at 10% level of significance. Therefore *educ* could be still endogenous and there could be some unobserved factors we have not included into our regression.

   (d) We can use the fitted value of *educ*, denoted as $\widehat{educ}$, from Equation

(4) to substitute out *educ* in Equation (3). We can then run an OLS regression to obtain $\hat{\epsilon}_{IV}$.

The LM statistic is $NR^2 = 852 \times 0.000163 \approx 0.13888$[1], which is smaller than the 5% cirtical value of a chi-square distribution with one degree of freedom (about 3.84146). Hence the instrument variables pass the overidentification test and they can be valid instruments.

6. (a) $\hat{\beta}_1 = -\frac{7}{11}$ and $\hat{\beta}_2 = \frac{6}{11}$.

   (b) $SSR = 3 - \frac{20}{11} = \frac{13}{11}$ and $SER = \sqrt{\frac{SSR}{10-3}} = \sqrt{\frac{13}{77}} \approx 0.41$.

[1] The SSR given in the exam paper should be R-squared.