LPS ECON-103-601, Fall 2014
HW Due 11.21.14 Name: ______________________

Show all work clearly and in order. Circle or box your final answer but points will be awarded based on a correct solution. A solution should always justify the steps taken and explain the assumptions needed to reach a final answer (e.g. how do you know you are not dividing by zero in the last step?).

For this homework, you are given the data $\{x_i, z_i, y_i\}_{i=1}^m$.

Q1

Suppose the true model is $y_i = z_i\gamma + \varepsilon_i$ where $\mathbb{E}[\varepsilon_i] = 0$ and $\varepsilon \sim i.i.d.$ over i. Derive the OLS estimator for γ (be sure to write out the OLS minimisation problem, write out the FOC, verify the SOC, and use the FOC to find $\hat{\gamma}_{OLS}$).

$\hat{\gamma}_{OLS}$ solves the following problem

$$\min_{\gamma} \sum_{i=1}^{m} (y_i - z_i\gamma)^2 = \min_{\gamma} S(\gamma)$$

$$\textbf{FOC}(\gamma) : \frac{\partial S}{\partial \gamma} = -2\sum_{i=1}^{m}(y_i - z_i\hat{\gamma})z_i = 0$$

Assuming $\sum\limits_{i=1}^{m} z_i^2 > 0$,

$$\Rightarrow \hat{\gamma}_{OLS} = \frac{\sum\limits_{i=1}^{m} y_i z_i}{\sum\limits_{i=1}^{m} z_i^2}$$

In ordering for $\hat{\gamma}$ to be the true minimiser, we must verify the SOC.

$$\textbf{SOC}(\gamma) : \frac{\partial^2 S}{\partial \gamma^2} = 2\sum_{i=1}^{m} z_i^2 > 0$$

The second derivative is positive, so S is convex and the FOC is necessary and sufficient to find the minimiser.

Q2

Show that $\hat{\gamma}_{OLS}$ is unbiased.

$$
\begin{aligned}
\hat{\gamma}_{OLS} &= \frac{\sum\limits_{i=1}^{m} y_i z_i}{\sum\limits_{i=1}^{m} z_i^2} \\
&= \frac{\sum\limits_{i=1}^{m} (z_i \gamma + \varepsilon_i) z_i}{\sum\limits_{i=1}^{m} z_i^2} \textbf{ using the true model} \\
&= \frac{\gamma \sum\limits_{i=1}^{m} z_i^2 + \sum\limits_{i=1}^{m} z_i \varepsilon_i}{\sum\limits_{i=1}^{m} z_i^2} \\
&= \gamma + \frac{\sum\limits_{i=1}^{m} z_i \varepsilon_i}{\sum\limits_{i=1}^{m} z_i^2} \\
\Rightarrow \hat{\gamma}_{OLS} - \gamma &= \frac{\sum\limits_{i=1}^{m} z_i \varepsilon_i}{\sum\limits_{i=1}^{m} z_i^2} \\
\mathbb{E}[\hat{\gamma}_{OLS} - \gamma] &= \mathbb{E}\left[\frac{\sum\limits_{i=1}^{m} z_i \varepsilon_i}{\sum\limits_{i=1}^{m} z_i^2}\right] \\
\mathbb{E}[\hat{\gamma}_{OLS} - \gamma] &= \frac{\sum\limits_{i=1}^{m} z_i \mathbb{E}\left[\varepsilon_i\right]}{\sum\limits_{i=1}^{m} z_i^2} \textbf{ assuming } z_i \textbf{ is fixed} \\
\mathbb{E}[\hat{\gamma}_{OLS} - \gamma] &= \frac{\sum\limits_{i=1}^{m} z_i \cdot 0}{\sum\limits_{i=1}^{m} z_i^2} \textbf{ given } \mathbb{E}[\varepsilon_i] = 0 \\
\mathbb{E}[\hat{\gamma}_{OLS} - \gamma] &= 0 \Rightarrow \mathbb{E}[\hat{\gamma}_{OLS}] = \gamma \textbf{ so } \hat{\gamma}_{OLS} \textbf{ is unbiased}
\end{aligned}
$$

Q3

Now derive $Var(\hat{\gamma}_{OLS})$ and determine whether $\hat{\gamma}_{OLS}$ is consistent (i.e. $Var(\hat{\gamma}_{OLS}) \to 0$ as $m \to \infty$).

$$
\begin{aligned}
Var(\hat{\gamma}_{OLS}) &= \mathbb{E}[(\hat{\gamma}_{OLS} - \gamma)^2] \\
&= \mathbb{E}\left[\left(\frac{\sum_{i=1}^{m} z_i\varepsilon_i}{\sum_{i=1}^{m} z_i^2}\right)^2\right] \\
&= \left(\sum_{i=1}^{m} z_i^2\right)^{-2} \cdot \mathbb{E}\left[\left(\sum_{i=1}^{m} z_i\varepsilon_i\right)^2\right] \textbf{ assuming } z_i \textbf{ is fixed} \\
&= \left(\sum_{i=1}^{m} z_i^2\right)^{-2} \cdot \mathbb{E}\left[\sum_{i=1}^{m} z_i^2\varepsilon_i^2 + 2\sum_{i=1}^{m}\sum_{j\neq i} z_i z_j \varepsilon_i \varepsilon_j\right] \\
&= \left(\sum_{i=1}^{m} z_i^2\right)^{-2} \cdot \left(\mathbb{E}\left[\sum_{i=1}^{m} z_i^2\varepsilon_i^2\right] + 2\mathbb{E}\left[\sum_{i=1}^{m}\sum_{j\neq i} z_i z_j \varepsilon_i \varepsilon_j\right]\right) \\
&= \left(\sum_{i=1}^{m} z_i^2\right)^{-2} \cdot \left(\sum_{i=1}^{m} z_i^2\mathbb{E}\left[\varepsilon_i^2\right] + 2\sum_{i=1}^{m}\sum_{j\neq i} z_i z_j \mathbb{E}\left[\varepsilon_i \varepsilon_j\right]\right) \textbf{ assuming } z_i \textbf{ is fixed} \\
&= \left(\sum_{i=1}^{m} z_i^2\right)^{-2} \cdot \left(\sum_{i=1}^{m} z_i^2\sigma_\varepsilon^2 + 2\sum_{i=1}^{m}\sum_{j\neq i} z_i z_j \cdot 0\right) \textbf{ given } \varepsilon_i \textbf{ is } i.i.d. \\
&= \left(\sum_{i=1}^{m} z_i^2\right)^{-2} \cdot \left(\sigma_\varepsilon^2 \sum_{i=1}^{m} z_i^2\right) \\
&= \sigma_\varepsilon^2 \left(\sum_{i=1}^{m} z_i^2\right)^{-2} \cdot \left(\sum_{i=1}^{m} z_i^2\right) \\
&= \sigma_\varepsilon^2 \left(\sum_{i=1}^{m} z_i^2\right)^{-1} \\
&= \frac{\sigma_\varepsilon^2}{m}\left(\underbrace{\frac{1}{m}\sum_{i=1}^{m} z_i^2}_{=\overline{z_i^2}}\right)^{-1} \\
&= \frac{\sigma_\varepsilon^2}{m}\frac{1}{\overline{z_i^2}} \textbf{ assuming } z_i \textbf{ is fixed} \\
\Rightarrow Var(\hat{\gamma}_{OLS}) &= \frac{\sigma_\varepsilon^2}{m}\frac{1}{\overline{z_i^2}} \to 0 \textbf{ as } m \to \infty
\end{aligned}
$$

Hence, $\hat{\gamma}_{OLS}$ is MSE-consistent and hence consistent.

Q4

Now suppose we were wrong and the true model is actually $y_i = z_i\gamma + x_i\beta + \varepsilon_i$ where $\mathbb{E}[\varepsilon_i] = 0$. Show that $\hat{\gamma}_{OLS}$ may be biased and derive an expression for the potential bias.

$$
\begin{aligned}
\hat{\gamma}_{OLS} &= \frac{\sum\limits_{i=1}^{m} y_i z_i}{\sum\limits_{i=1}^{m} z_i^2} \\
&= \frac{\sum\limits_{i=1}^{m} (z_i\gamma + x_i\beta + \varepsilon_i) z_i}{\sum\limits_{i=1}^{m} z_i^2} \textbf{ using the true model} \\
&= \frac{\sum\limits_{i=1}^{m} (z_i\gamma + x_i\beta + \varepsilon_i) z_i}{\sum\limits_{i=1}^{m} z_i^2} \\
&= \frac{\sum\limits_{i=1}^{m} (z_i^2\gamma + z_i x_i\beta + z_i\varepsilon_i)}{\sum\limits_{i=1}^{m} z_i^2} \\
&= \gamma + \beta \cdot \frac{\sum\limits_{i=1}^{m} z_i x_i}{\sum\limits_{i=1}^{m} z_i^2} + \frac{\sum\limits_{i=1}^{m} z_i \varepsilon_i}{\sum\limits_{i=1}^{m} z_i^2} \\
\mathbb{E}[\hat{\gamma}_{OLS} - \gamma] &= \mathbb{E}\left[\beta \cdot \frac{\sum\limits_{i=1}^{m} z_i x_i}{\sum\limits_{i=1}^{m} z_i^2} + \frac{\sum\limits_{i=1}^{m} z_i \varepsilon_i}{\sum\limits_{i=1}^{m} z_i^2}\right] \\
&= \beta \cdot \frac{\sum\limits_{i=1}^{m} z_i x_i}{\sum\limits_{i=1}^{m} z_i^2} + \frac{\sum\limits_{i=1}^{m} z_i \mathbb{E}\left[\varepsilon_i\right]}{\sum\limits_{i=1}^{m} z_i^2} \textbf{ assuming } x_i \textbf{ and } z_i \textbf{ are fixed} \\
&= \beta \cdot \frac{\sum\limits_{i=1}^{m} z_i x_i}{\sum\limits_{i=1}^{m} z_i^2} + \frac{\sum\limits_{i=1}^{m} z_i \cdot 0}{\sum\limits_{i=1}^{m} z_i^2} \textbf{ given } \mathbb{E}[\varepsilon_i] = 0 \\
&= \beta \cdot \frac{\sum\limits_{i=1}^{m} z_i x_i}{\sum\limits_{i=1}^{m} z_i^2} \\
&= \beta \cdot \frac{\frac{1}{n}\sum\limits_{i=1}^{m} z_i x_i}{\frac{1}{n}\sum\limits_{i=1}^{m} z_i^2}
\end{aligned}
$$

Hence, the potential bias is $\beta \cdot \frac{\frac{1}{n}\sum_{i=1}^{m} z_i x_i}{\frac{1}{n}\sum_{i=1}^{m} z_i^2}$**.**

Q5

Under what condition will $\hat{\gamma}_{OLS}$ be unbiased?

There are two conditions under which $\hat{\gamma}_{OLS}$ **will be unbiased.**

- $\beta = 0$**, this restricts to the true model to be that of Q1**
- $\frac{1}{n}\sum_{i=1}^{m} z_i x_i = 0$**, this restricts** x **and** z **to be uncorrelated (letting either** $\bar{z}$ **or** $\bar{x}$ **be 0)**

If neither of these conditions hold, then OLS is subject to what we call omitted variable bias. Not only will OLS be biased, but it will not be consistent either (so having more data will not help because the model is wrong).