Working paper.

TECHNICAL REPORT
R-388
September 2011

Precision of Composite Estimators

Jinyong Hahn
UCLA

Judea Pearl
UCLA

September 23, 2011

Abstract

We assess the precision of direct vs. indirect methods of estimating regression parameter. In particular we compare the direct estimator, defined by the regression itself, with *composite* estimators, which invoke auxiliary variables, under various modeling assumptions of exclusion and indepndence. Our general conclusion is that a composite estimator that exploits the model restrictions has greater asymptotic precision than its direct counterpart.

1 Model 1

Consider the following model:

- y is the outcome of interest
- x is the intermediate cause
- z is the cause of interest
- z affects y only through x

Example 1 *y denotes "heart attack", x denotes "cholesterol", and z denotes "butter"*

Let's give some more mathematical structure:

$$y = \beta x + u \tag{1}$$

$$x = \gamma z + v \tag{2}$$

and

$$\begin{aligned} y &= \beta(\gamma z + v) + u \\ &= \beta\gamma z + (\beta v + u) \end{aligned}$$

We will define $\theta = \beta\gamma$ and $\varepsilon = \beta v + u$, and write

$$y = \theta z + \varepsilon \tag{3}$$

For simplicity, we will assume that $(y, x, z)'$ have the multivariate normal distribution with zero mean.

Two Estimators We consider two estimation methods of θ. The first one estimates it in one step using (3). The second one estimates (β, γ) using (1) and (2), and then use the relationship $\theta = \beta\gamma$ to compute an estimate of θ.

The single step estimator We will call the single step estimator $\widetilde{\theta}$. Using the well-known result, we can see that

$$\sqrt{n}\left(\widetilde{\theta} - \theta\right) \to N\left(0, \frac{\operatorname{Var}(\varepsilon)}{\operatorname{Var}(z)}\right)$$

Because $\operatorname{Cov}(u, v) = 0$, we have $\operatorname{Var}(\varepsilon) = \beta^2 \operatorname{Var}(v) + \operatorname{Var}(u)$, which implies that we can write

$$\sqrt{n}\left(\widetilde{\theta} - \theta\right) \to N\left(0, \frac{\beta^2 \operatorname{Var}(v) + \operatorname{Var}(u)}{\operatorname{Var}(z)}\right)$$

The two step estimator We will call the single step estimator $\widehat{\theta}$. Again using the well-known result, we can see that

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \to N\left(0, \frac{\operatorname{Var}(u)}{\operatorname{Var}(x)}\right)$$

Using $\operatorname{Var}(x) = \gamma^2 \operatorname{Var}(z) + \operatorname{Var}(v)$, we obtain

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \to N\left(0, \frac{\operatorname{Var}(u)}{\gamma^2 \operatorname{Var}(z) + \operatorname{Var}(v)}\right)$$

We now claim that $\sqrt{n}\left(\widehat{\beta} - \beta\right)$ and $\sqrt{n}\left(\widehat{\gamma} - \gamma\right)$ are asymptotically independent. We note that

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i}{\operatorname{Var}(x)} + o_p(1)$$
$$\sqrt{n}\left(\widehat{\gamma} - \gamma\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i v_i}{\operatorname{Var}(z)} + o_p(1)$$

and

$$E\left[(xu)(zv)\right] = E\left[z(\gamma z + v)uv\right] = \gamma E\left[z^2 uv\right] + E\left[zv^2 u\right]$$

With normality along with $\operatorname{Cov}(u, v) = 0$, this component is equal to zero. Therefore, $\widehat{\beta}$ and $\widehat{\gamma}$ are asymptotically independent.

Now, using the delta method, we have $\sqrt{n}\left(\widehat{\theta} - \theta\right)$ asymptotically normal with asymptotic variance equal to

$$\frac{\gamma^2 \operatorname{Var}(u)}{\gamma^2 \operatorname{Var}(z) + \operatorname{Var}(v)} + \frac{\beta^2 \operatorname{Var}(v)}{\operatorname{Var}(z)}$$

Comparison We see that the single step estimator has the asymptotic variance equal to

$$\frac{\beta^2 \operatorname{Var}(v) + \operatorname{Var}(u)}{\operatorname{Var}(z)} \tag{4}$$

and the two step estimator has the asymptotic variance equal to

$$\frac{\gamma^2 \operatorname{Var}(u)}{\gamma^2 \operatorname{Var}(z) + \operatorname{Var}(v)} + \frac{\beta^2 \operatorname{Var}(v)}{\operatorname{Var}(z)} \tag{5}$$

Subtracting (5) from (4), we obtain

$$\begin{aligned}
&(4) - (5) \\
&= \frac{\beta^2 \operatorname{Var}(v) + \operatorname{Var}(u)}{\operatorname{Var}(z)} - \left(\frac{\gamma^2 \operatorname{Var}(u)}{\gamma^2 \operatorname{Var}(z) + \operatorname{Var}(v)} + \frac{\beta^2 \operatorname{Var}(v)}{\operatorname{Var}(z)} \right) \\
&= \frac{\operatorname{Var}(u)}{\operatorname{Var}(z)} - \frac{\operatorname{Var}(u)}{\operatorname{Var}(z) + \frac{1}{\gamma^2} \operatorname{Var}(v)} > 0
\end{aligned}$$

In other words, the two-step estimator is more efficient than the one step estimator.

2 Model 2

Consider the following model:

- y is the outcome of interest
- x is the intermediate cause
- z is the cause of interest
- z affects y directly, and indirectly through x

Example 2 *y denotes "heart attack", x denotes "cholesterol", and z denotes "butter". We will assume that the "butter" has both direct and indirect impacts, which is expressed (6) in mathematical terms.*

Let's give some more mathematical structure:

$$y = \beta x + \delta z + u \tag{6}$$
$$x = \gamma z + v \tag{7}$$

and

$$\begin{aligned}
y &= \beta (\gamma z + v) + \delta z + u \\
&= (\beta\gamma + \delta) z + (\beta v + u)
\end{aligned}$$

We will define $\theta = \beta\gamma + \delta$ and $\varepsilon = \beta v + u$, and write

$$y = \theta z + \varepsilon \tag{8}$$

For simplicity, we will assume that $(y, x, z)'$ have the multivariate normal distribution with zero mean.

Two Estimators We consider two estimation methods of θ. The first one estimates it in one step using (8). The second one estimates (β,γ) using (6) and (7), and then use the relationship $\theta=\beta\gamma+\delta$ to compute an estimate of θ.

The single step estimator We will call the single step estimator $\widetilde{\theta}$. Using the well-known result, we can see that

$$\sqrt{n}\left(\widetilde{\theta}-\theta\right)\to N\left(0,\frac{\operatorname{Var}(\varepsilon)}{\operatorname{Var}(z)}\right)=N\left(0,\frac{\beta^2\operatorname{Var}(v)+\operatorname{Var}(u)}{\operatorname{Var}(z)}\right)$$

Remark 1 *If the OLS on (6) is to be consistent, we should have* $\operatorname{Cov}(u,v)=0$. *This means that*

$$\operatorname{Var}(\varepsilon)=\beta^2\operatorname{Var}(v)+\operatorname{Var}(u)$$

It follows that

$$\frac{\operatorname{Var}(\varepsilon)}{\operatorname{Var}(z)}=\frac{\beta^2\operatorname{Var}(v)+\operatorname{Var}(u)}{\operatorname{Var}(z)}$$

The two step estimator We will call the two step estimator $\widehat{\theta}$. Again using the well-known result, we can see that

$$\sqrt{n}\left(\begin{bmatrix}\widehat{\beta}\\ \widehat{\delta}\end{bmatrix}-\begin{bmatrix}\beta\\ \delta\end{bmatrix}\right)\to N\left(\begin{bmatrix}0\\ 0\end{bmatrix},\operatorname{Var}(u)\begin{bmatrix}\operatorname{Var}(x) & \operatorname{Cov}(x,z)\\ \operatorname{Cov}(x,z) & \operatorname{Var}(z)\end{bmatrix}^{-1}\right)$$

and

$$\sqrt{n}\left(\widehat{\gamma}-\gamma\right)\to N\left(0,\frac{\operatorname{Var}(v)}{\operatorname{Var}(z)}\right)$$

They are asymptotically independent by the same reasoning as outlined in the previous note. By the analysis of variance formula in (7), we also have

$$\begin{aligned}\operatorname{Var}(x)&=\gamma^2\operatorname{Var}(z)+\operatorname{Var}(v)\\ \operatorname{Cov}(x,z)&=\gamma\operatorname{Var}(z)\end{aligned}$$

It follows that the asymptotic variance of $\left(\widehat{\beta},\widehat{\delta}\right)'$ is equal to

$$\begin{aligned}\operatorname{Var}(u)\begin{bmatrix}\operatorname{Var}(x) & \operatorname{Cov}(x,z)\\ \operatorname{Cov}(x,z) & \operatorname{Var}(z)\end{bmatrix}^{-1}&=\operatorname{Var}(u)\begin{bmatrix}\gamma^2\operatorname{Var}(z)+\operatorname{Var}(v) & \gamma\operatorname{Var}(z)\\ \gamma\operatorname{Var}(z) & \operatorname{Var}(z)\end{bmatrix}^{-1}\\ &=\frac{\operatorname{Var}(u)}{\operatorname{Var}(v)}\begin{bmatrix}1 & -\gamma\\ -\gamma & \gamma^2+\frac{\operatorname{Var}(v)}{\operatorname{Var}(z)}\end{bmatrix}\end{aligned}$$

To conclude, the asymptotic variance of $\left(\widehat{\beta},\widehat{\delta},\widehat{\gamma}\right)'$ is equal to

$$\begin{bmatrix}\frac{\operatorname{Var}(u)}{\operatorname{Var}(v)} & -\gamma\frac{\operatorname{Var}(u)}{\operatorname{Var}(v)} & 0\\ -\gamma\frac{\operatorname{Var}(u)}{\operatorname{Var}(v)} & \gamma^2\frac{\operatorname{Var}(u)}{\operatorname{Var}(v)}+\frac{\operatorname{Var}(u)}{\operatorname{Var}(z)} & 0\\ 0 & 0 & \frac{\operatorname{Var}(v)}{\operatorname{Var}(z)}\end{bmatrix}$$

Now, using the delta method, we have $\sqrt{n}\left(\widehat{\theta}-\theta\right)$ asymptotically normal with asymptotic variance equal to

$$\begin{aligned}
&\left[\begin{array}{ccc}\gamma & 1 & \beta\end{array}\right]\left[\begin{array}{ccc}\frac{\mathrm{Var}(u)}{\mathrm{Var}(v)} & -\gamma\frac{\mathrm{Var}(u)}{\mathrm{Var}(v)} & 0 \\ -\gamma\frac{\mathrm{Var}(u)}{\mathrm{Var}(v)} & \gamma^2\frac{\mathrm{Var}(u)}{\mathrm{Var}(v)}+\frac{\mathrm{Var}(u)}{\mathrm{Var}(z)} & 0 \\ 0 & 0 & \frac{\mathrm{Var}(v)}{\mathrm{Var}(z)}\end{array}\right]\left[\begin{array}{c}\gamma \\ 1 \\ \beta\end{array}\right] \\
&= \frac{\beta^2\,\mathrm{Var}\,(v)+\mathrm{Var}\,(u)}{\mathrm{Var}\,(z)}
\end{aligned}$$

Comparison We see that the single step estimator has the asymptotic variance equal to

$$\frac{\beta^2\,\mathrm{Var}\,(v)+\mathrm{Var}\,(u)}{\mathrm{Var}\,(z)}$$

and the two step estimator has the asymptotic variance equal to

$$\frac{\beta^2\,\mathrm{Var}\,(v)+\mathrm{Var}\,(u)}{\mathrm{Var}\,(z)}$$

They are identical!

Explanation The single step estimator should be equal to

$$\widetilde{\theta}=\frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}$$

As for the two step estimator $\widehat{\theta}$, we note that

$$y_i=\widehat{\beta}x_i+\widehat{\delta}z_i+\widehat{u}_i$$

with

$$\begin{aligned}
\left[\begin{array}{c}\widehat{\beta} \\ \widehat{\delta}\end{array}\right] &= \left[\begin{array}{cc}\sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i z_i \\ \sum_{i=1}^n x_i z_i & \sum_{i=1}^n z_i^2\end{array}\right]^{-1}\left[\begin{array}{c}\sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n z_i y_i\end{array}\right] \\
&= \frac{1}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right)-\left(\sum_{i=1}^n x_i z_i\right)^2}\left[\begin{array}{c}\left(\sum_{i=1}^n z_i^2\right)\left(\sum_{i=1}^n x_i y_i\right)-\left(\sum_{i=1}^n x_i z_i\right)\left(\sum_{i=1}^n z_i y_i\right) \\ -\left(\sum_{i=1}^n x_i z_i\right)\left(\sum_{i=1}^n x_i y_i\right)+\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i y_i\right)\end{array}\right]
\end{aligned}$$

and

$$\widehat{\gamma}=\frac{\sum_{i=1}^n x_i z_i}{\sum_{i=1}^n z_i^2}$$

Therefore, we have

$$
\begin{aligned}
\widehat{\theta} =& \widehat{\beta}\widehat{\gamma} + \widehat{\delta} \\
=& \frac{\left(\sum_{i=1}^n z_i^2\right)\left(\sum_{i=1}^n x_i y_i\right) - \left(\sum_{i=1}^n x_i z_i\right)\left(\sum_{i=1}^n z_i y_i\right)}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2} \frac{\sum_{i=1}^n x_i z_i}{\sum_{i=1}^n z_i^2} \\
&+ \frac{-\left(\sum_{i=1}^n x_i z_i\right)\left(\sum_{i=1}^n x_i y_i\right) + \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i y_i\right)}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2} \\
=& \frac{\left(\sum_{i=1}^n x_i y_i\right)\left(\sum_{i=1}^n x_i z_i\right) - \left(\sum_{i=1}^n x_i z_i\right)^2 \left(\frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}\right)}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2} \\
&+ \frac{-\left(\sum_{i=1}^n x_i z_i\right)\left(\sum_{i=1}^n x_i y_i\right) + \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i y_i\right)}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2} \\
=& \frac{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i y_i\right) - \left(\sum_{i=1}^n x_i z_i\right)^2 \left(\frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}\right)}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2} \\
=& \frac{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right)\left(\frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}\right) - \left(\sum_{i=1}^n x_i z_i\right)^2 \left(\frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}\right)}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2} \\
=& \frac{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n z_i^2\right) - \left(\sum_{i=1}^n x_i z_i\right)^2} \left(\frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2}\right) \\
=& \frac{\sum_{i=1}^n z_i y_i}{\sum_{i=1}^n z_i^2} \\
=& \widetilde{\theta}
\end{aligned}
$$

Given that the two estimators are numerically identical, it should not surprise us that their asymptotic variances are identical.

3 Model 3

We will assume

$$
\begin{aligned}
y &= \alpha x + \beta w + \varepsilon_1 \\
w &= \gamma z + \varepsilon_2 \\
z &= \delta x + \varepsilon_3
\end{aligned}
$$

We assume that all the equations can be estimated by OLS. This implies certain orthogonality. For simplification of derivation, we will assume that orthogonality is equivalent to independence. We will also assume that all the variables have zero mean. Under these assumptions, we obtain the following independences:

$$\varepsilon_3 \perp\!\!\!\perp x$$
$$\varepsilon_2 \perp\!\!\!\perp z$$
$$\varepsilon_1 \perp\!\!\!\perp (x, w)$$

It is straightforward to show that it also implies that

$$\varepsilon_3 \perp\!\!\!\perp \varepsilon_2 \tag{9}$$
$$\varepsilon_1 \perp\!\!\!\perp \varepsilon_2 \tag{10}$$

Remark 2 *It is not clear whether we obtain the implication*

$$\varepsilon_3 \perp\!\!\!\perp \varepsilon_1 \tag{11}$$

as well, which would simplify the derivation below. Probably not. In order to make sure that we do, we will make the assumption that the equation

$$y = \alpha x + \beta w + \varepsilon_1$$

means that, if we regress y on (x, w, z), then we get 0 as the coefficient of z. This would ensure that we also get (11).

Parameter of Interest We are interested in estimation of α. Below are some of the natural estimators.

1. We regress y on (x, z)
2. We regress y on (x, w)
3. We regress y on (x, w, z)
4. We regress y on x, and subtract it by the product $\beta\gamma\delta$ identified from the three equations

$$\begin{aligned} y &= \alpha x + \beta w + \varepsilon_1 \\ w &= \gamma z + \varepsilon_2 \\ z &= \delta x + \varepsilon_3 \end{aligned}$$

We show that the second estimator has the smallest asymptotic variance.

First estimator We note that

$$\begin{aligned} y &= \alpha x + \beta w + \varepsilon_1 \\ &= \alpha x + \beta (\gamma z + \varepsilon_2) + \varepsilon_1 \\ &= \alpha x + \beta\gamma z + (\beta\varepsilon_2 + \varepsilon_1) \end{aligned}$$

and regress y on (x, z). The asymptotic variance matrix of the two-dimensional coefficient vector is given by

$$\operatorname{Var}(\beta\varepsilon_2 + \varepsilon_1)\begin{bmatrix} \sigma_x^2 & \sigma_{xz} \\ \sigma_{xz} & \sigma_z^2 \end{bmatrix}^{-1}$$

where we adopt the convention of using σ^2 to denote a generic variance. We also let $\sigma_j^2 = \operatorname{Var}(\varepsilon_j)$. Noting that

$$\begin{aligned} \sigma_{xz} &= \delta\sigma_x^2 \\ \sigma_z^2 &= \delta^2\sigma_x^2 + \sigma_3^2 \\ \operatorname{Var}(\beta\varepsilon_2 + \varepsilon_1) &= \beta^2\sigma_2^2 + \sigma_1^2 \end{aligned}$$

we obtain that the asymptotic variance matrix is

$$\left(\beta^2\sigma_2^2 + \sigma_1^2\right)\begin{bmatrix} \sigma_x^2 & \delta\sigma_x^2 \\ \delta\sigma_x^2 & \delta^2\sigma_x^2 + \sigma_3^2 \end{bmatrix}^{-1}$$

and the asymptotic variance of the estimator of α is given by the (1,1)-element of the matrix above, i.e.,

$$\left(\beta^2\sigma_2^2 + \sigma_1^2\right)\frac{1}{\sigma_3^2\sigma_x^2}\left(\delta^2\sigma_x^2 + \sigma_3^2\right) \tag{12}$$

Second estimator The second estimator simply uses the equation

$$y = \alpha x + \beta w + \varepsilon_1$$

and regress y on (x, w). The asymptotic variance matrix of the two-dimensional coefficient vector is given by

$$\operatorname{Var}(\varepsilon_1)\begin{bmatrix} \sigma_x^2 & \sigma_{xw} \\ \sigma_{xw} & \sigma_w^2 \end{bmatrix}^{-1}$$

Noting that

$$\begin{aligned} w &= \gamma z + \varepsilon_2 \\ &= \gamma(\delta x + \varepsilon_3) + \varepsilon_2 \\ &= \gamma\delta x + (\gamma\varepsilon_3 + \varepsilon_2) \end{aligned}$$

with the implication

$$\begin{aligned} \sigma_{xw} &= \gamma\delta\sigma_x^2 \\ \sigma_w^2 &= \gamma^2\delta^2\sigma_x^2 + \gamma^2\sigma_3^2 + \sigma_2^2 \end{aligned}$$

we obtain the asymptotic variance matrix

$$\operatorname{Var}(\varepsilon_1)\begin{bmatrix} \sigma_x^2 & \sigma_{xw} \\ \sigma_{xw} & \sigma_w^2 \end{bmatrix}^{-1} = \sigma_1^2\begin{bmatrix} \sigma_x^2 & \gamma\delta\sigma_x^2 \\ \gamma\delta\sigma_x^2 & \gamma^2\delta^2\sigma_x^2 + \gamma^2\sigma_3^2 + \sigma_2^2 \end{bmatrix}^{-1}$$

and the asymptotic variance of the estimator of α is given by the (1,1)-element

$$\frac{\sigma_1^2}{\gamma^2\sigma_3^2\sigma_x^2+\sigma_2^2\sigma_x^2}\left(\gamma^2\delta^2\sigma_x^2+\gamma^2\sigma_3^2+\sigma_2^2\right) \tag{13}$$

Subtracting (13) from (12), we obtain

$$\begin{aligned}
&\left(\beta^2\sigma_2^2+\sigma_1^2\right)\frac{1}{\sigma_3^2\sigma_x^2}\left(\delta^2\sigma_x^2+\sigma_3^2\right)-\frac{\sigma_1^2}{\gamma^2\sigma_3^2\sigma_x^2+\sigma_2^2\sigma_x^2}\left(\gamma^2\delta^2\sigma_x^2+\gamma^2\sigma_3^2+\sigma_2^2\right)\\
&\qquad=\frac{\sigma_2^2}{\sigma_3^2\sigma_x^2\left(\gamma^2\sigma_3^2+\sigma_2^2\right)}\left(\beta^2\gamma^2\delta^2\sigma_3^2\sigma_x^2+\beta^2\gamma^2\sigma_3^4+\beta^2\delta^2\sigma_2^2\sigma_x^2+\beta^2\sigma_2^2\sigma_3^2+\delta^2\sigma_1^2\sigma_x^2\right)
\end{aligned}$$

Every element on the right is positive, so we conclude that the second estimator has the smaller asymptotic variance than the first.

Third estimator The third estimator uses

$$y=\alpha x+\beta w+0\times z+\varepsilon_1$$

and regress y on (x,w,z). The asymptotic variance matrix of the three-dimensional coefficient vector is given by

$$\operatorname{Var}\left(\varepsilon_1\right)\begin{bmatrix}\sigma_x^2 & \sigma_{xw} & \sigma_{xz}\\ \sigma_{xw} & \sigma_w^2 & \sigma_{wz}\\ \sigma_{xz} & \sigma_{wz} & \sigma_z^2\end{bmatrix}^{-1}$$

We have already established that

$$\begin{aligned}
\sigma_{xz}&=\delta\sigma_x^2\\
\sigma_z^2&=\delta^2\sigma_x^2+\sigma_3^2\\
\sigma_{xw}&=\gamma\delta\sigma_x^2\\
\sigma_w^2&=\gamma^2\delta^2\sigma_x^2+\gamma^2\sigma_3^2+\sigma_2^2
\end{aligned}$$

Using

$$w=\gamma z+\varepsilon_2$$

we also establish

$$\sigma_{wz}=\gamma\sigma_z^2=\gamma\left(\delta^2\sigma_x^2+\sigma_3^2\right)=\gamma\delta^2\sigma_x^2+\gamma\sigma_3^2$$

Therefore, the asymptotic variance matrix is

$$\sigma_1^2\begin{bmatrix}\sigma_x^2 & \gamma\delta\sigma_x^2 & \delta\sigma_x^2\\ \gamma\delta\sigma_x^2 & \gamma^2\delta^2\sigma_x^2+\gamma^2\sigma_3^2+\sigma_2^2 & \gamma\delta^2\sigma_x^2+\gamma\sigma_3^2\\ \delta\sigma_x^2 & \gamma\delta^2\sigma_x^2+\gamma\sigma_3^2 & \delta^2\sigma_x^2+\sigma_3^2\end{bmatrix}^{-1}$$

and the asymptotic variance of the estimator of α is given by the (1,1)-element

$$\frac{\sigma_1^2}{\sigma_3^2\sigma_x^2}\left(\delta^2\sigma_x^2+\sigma_3^2\right) \tag{14}$$

Subtracting (13) from (14), we obtain

$$\frac{\sigma_1^2}{\sigma_3^2\sigma_x^2}\left(\delta^2\sigma_x^2+\sigma_3^2\right)-\frac{\sigma_1^2}{\gamma^2\sigma_3^2\sigma_x^2+\sigma_2^2\sigma_x^2}\left(\gamma^2\delta^2\sigma_x^2+\gamma^2\sigma_3^2+\sigma_2^2\right)=\delta^2\sigma_1^2\frac{\sigma_2^2}{\sigma_3^2\left(\gamma^2\sigma_3^2+\sigma_2^2\right)}>0$$

so we conclude that the second estimator has the smaller asymptotic variance than the third.

Fourth estimator The fourth estimator notes that

$$\begin{aligned} y &= \alpha x + \beta w + \varepsilon_1 = \alpha x + \beta\left(\gamma z + \varepsilon_2\right) + \varepsilon_1 \\ &= \alpha x + \beta\gamma z + \beta\varepsilon_2 + \varepsilon_1 = \alpha x + \beta\gamma\left(\delta x + \varepsilon_3\right) + \beta\varepsilon_2 + \varepsilon_1 \\ &= \left(\alpha + \beta\gamma\delta\right) x + \left(\beta\gamma\varepsilon_3 + \beta\varepsilon_2 + \varepsilon_1\right) \end{aligned}$$

Because β is identified from

$$y = \alpha x + \beta w + \varepsilon_1$$

and γ and δ are identified from

$$\begin{aligned} w &= \gamma z + \varepsilon_2 \\ z &= \delta x + \varepsilon_3 \end{aligned}$$

we can estimate α by regressing y on x and subtracting it by the product $\beta\gamma\delta$ identified from the three equations.

In order to make sense of them, we need the asymptotic covariance of every estimator. For this purpose, we set up the moment equation

$$E\begin{bmatrix} x\left(y - \varphi x\right) \\ x\left(y - \left(\alpha x + \beta w\right)\right) \\ w\left(y - \left(\alpha x + \beta w\right)\right) \\ z\left(w - \gamma z\right) \\ x\left(z - \delta x\right) \end{bmatrix} = 0$$

where we define

$$\varphi = \alpha + \beta\gamma\delta$$

This is an exactly identified moment, and the asymptotic variance requires characterization of the variance matrix of

$$\begin{bmatrix} x\left(y - \varphi x\right) \\ x\left(y - \left(\alpha x + \beta w\right)\right) \\ w\left(y - \left(\alpha x + \beta w\right)\right) \\ z\left(w - \gamma z\right) \\ x\left(z - \delta x\right) \end{bmatrix} = \begin{bmatrix} x\left(\beta\gamma\varepsilon_3 + \beta\varepsilon_2 + \varepsilon_1\right) \\ x\varepsilon_1 \\ w\varepsilon_1 \\ z\varepsilon_2 \\ x\varepsilon_3 \end{bmatrix}$$

at the truth. Writing

$$\begin{bmatrix} x\left(\beta\gamma\varepsilon_3 + \beta\varepsilon_2 + \varepsilon_1\right) \\ x\varepsilon_1 \\ w\varepsilon_1 \\ z\varepsilon_2 \\ x\varepsilon_3 \end{bmatrix} = \begin{bmatrix} x\left(\beta\gamma\varepsilon_3 + \beta\varepsilon_2 + \varepsilon_1\right) \\ x\varepsilon_1 \\ \left(\gamma\delta x + \left(\gamma\varepsilon_3 + \varepsilon_2\right)\right)\varepsilon_1 \\ \left(\delta x + \varepsilon_3\right)\varepsilon_2 \\ x\varepsilon_3 \end{bmatrix} = \begin{bmatrix} x\varepsilon_1 + \beta x\varepsilon_2 + \beta\gamma x\varepsilon_3 \\ x\varepsilon_1 \\ \varepsilon_1\varepsilon_2 + \gamma\varepsilon_1\varepsilon_3 + \gamma\delta x\varepsilon_1 \\ \varepsilon_2\varepsilon_3 + \delta x\varepsilon_2 \\ x\varepsilon_3 \end{bmatrix}$$

we can see that the variance matrix of interest is

$$\Omega \equiv \begin{bmatrix} \sigma_x^2\sigma_1^2 + \beta^2\sigma_x^2\sigma_2^2 + \beta^2\gamma^2\sigma_x^2\sigma_3^2 & \sigma_x^2\sigma_1^2 & \gamma\delta\sigma_x^2\sigma_1^2 & \beta\delta\sigma_x^2\sigma_2^2 & \beta\gamma\sigma_x^2\sigma_3^2 \\ \sigma_x^2\sigma_1^2 & \sigma_x^2\sigma_1^2 & \gamma\delta\sigma_x^2\sigma_1^2 & 0 & 0 \\ \gamma\delta\sigma_x^2\sigma_1^2 & \gamma\delta\sigma_x^2\sigma_1^2 & \sigma_1^2\sigma_2^2 + \gamma^2\sigma_1^2\sigma_3^2 + \gamma^2\delta^2\sigma_x^2\sigma_1^2 & 0 & 0 \\ \beta\delta\sigma_x^2\sigma_2^2 & 0 & 0 & \sigma_2^2\sigma_3^2 + \delta^2\sigma_x^2\sigma_2^2 & 0 \\ \beta\gamma\sigma_x^2\sigma_3^2 & 0 & 0 & 0 & \sigma_x^2\sigma_3^2 \end{bmatrix}$$

Remark 3 *There is a fair amount of simplification using independence among ε's.*

With

$$\Upsilon \equiv -E\left[\frac{\partial}{\partial\theta'}\begin{bmatrix} x(y-\varphi x) \\ x(y-(\alpha x+\beta w)) \\ w(y-(\alpha x+\beta w)) \\ z(w-\gamma z) \\ x(z-\delta x) \end{bmatrix}\right] = E\begin{bmatrix} x^2 & 0 & 0 & 0 & 0 \\ 0 & x^2 & xw & 0 & 0 \\ 0 & xw & w^2 & 0 & 0 \\ 0 & 0 & 0 & z^2 & 0 \\ 0 & 0 & 0 & 0 & x^2 \end{bmatrix}$$
$$= \begin{bmatrix} \sigma_x^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_x^2 & \gamma\delta\sigma_x^2 & 0 & 0 \\ 0 & \gamma\delta\sigma_x^2 & \gamma^2\delta^2\sigma_x^2+\gamma^2\sigma_3^2+\sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & \delta^2\sigma_x^2+\sigma_3^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_x^2 \end{bmatrix}$$

the asymptotic variance of the estimator of θ is given by $\Upsilon^{-1}\Omega\left(\Upsilon^{-1}\right)'$.

As for the estimator of α, we note that it is based on the equality

$$\alpha = \varphi - \beta\gamma\delta$$

and apply the delta method, which delivers

$$\frac{1}{\sigma_x^2\left(\gamma^2\sigma_3^2+\sigma_2^2\right)\left(\delta^2\sigma_x^2+\sigma_3^2\right)}\left(\beta^2\gamma^2\sigma_2^2\sigma_3^4+\beta^2\sigma_2^4\sigma_3^2+\gamma^2\delta^4\sigma_1^2\sigma_x^4+2\gamma^2\delta^2\sigma_1^2\sigma_3^2\sigma_x^2+\gamma^2\sigma_1^2\sigma_3^4+\delta^2\sigma_1^2\sigma_2^2\sigma_x^2+\sigma_1^2\sigma_2^2\sigma_3^2\right) \tag{15}$$

as the asymptotic variance. Subtracting (13) from (15), we obtain

$$\frac{\beta^2\gamma^2\sigma_2^2\sigma_3^4+\beta^2\sigma_2^4\sigma_3^2+\gamma^2\delta^4\sigma_1^2\sigma_x^4+2\gamma^2\delta^2\sigma_1^2\sigma_3^2\sigma_x^2+\gamma^2\sigma_1^2\sigma_3^4+\delta^2\sigma_1^2\sigma_2^2\sigma_x^2+\sigma_1^2\sigma_2^2\sigma_3^2}{\sigma_x^2\left(\gamma^2\sigma_3^2+\sigma_2^2\right)\left(\delta^2\sigma_x^2+\sigma_3^2\right)}$$
$$-\frac{\sigma_1^2}{\gamma^2\sigma_3^2\sigma_x^2+\sigma_2^2\sigma_x^2}\left(\gamma^2\delta^2\sigma_x^2+\gamma^2\sigma_3^2+\sigma_2^2\right) = \beta^2\sigma_2^2\frac{\sigma_3^2}{\sigma_x^2\left(\delta^2\sigma_x^2+\sigma_3^2\right)} > 0$$

so we conclude that the second estimator has the smaller asymptotic variance than the fourth.

4 Model 4

We will assume the following:

- z, x, y are binary
- $z \to x \to y$
- Our objective of interest is

$$\begin{aligned} p &= P\left[y=1|\, z=1\right] \\ &= P\left[y=1|\, x=1\right]P\left[x=1|\, z=1\right]+P\left[y=1|\, x=0\right]P\left[x=0|\, z=1\right] \end{aligned}$$

- We would like to compare the one shot estimator $\widehat{P}\left[y=1 \mid z=1\right]$ against the compositional estimator $\widehat{P}\left[y=1 \mid x=1\right] \widehat{P}\left[x=1 \mid z=1\right]+\widehat{P}\left[y=1 \mid x=0\right] \widehat{P}\left[x=0 \mid z=1\right]$

- We show that the asymptotic variance of the one shot estimator is larger than that of the compositional estimator.

4.1 Technical Derivation

Some symbols

- $\alpha=\Pr\left[y=1 \mid x=1\right]$

- $\beta=\Pr\left[y=1 \mid x=0\right]$

- $\gamma=\Pr\left[x=1, z=1\right]$

- $\delta=\Pr\left[x=1, z=0\right]$

- $\varphi=\Pr\left[x=0, z=1\right]$

- $\psi=\Pr\left[x=0, z=0\right]$

Remark 4 *We obviously have $\psi=1-\gamma-\delta-\varphi$, but it is more convenient to have this for programming.'*

With these symbols, we can say that the target parameter of interest is

$$\begin{aligned}\theta &= \Pr\left[y=1 \mid z=1\right]\\ &= \Pr\left[y=1 \mid x=1\right] \Pr\left[x=1 \mid z=1\right]+\Pr\left[y=1 \mid x=0\right] \Pr\left[x=0 \mid z=1\right]\\ &= \alpha \frac{\gamma}{\gamma+\varphi}+\beta \frac{\varphi}{\gamma+\varphi}\\ &= \alpha \xi+\beta\left(1-\xi\right)\end{aligned}$$

where we let

$$\xi=\frac{\gamma}{\gamma+\varphi}=\Pr\left[x=1 \mid z=1\right]$$

Structure in the model The structure in the model is such that z affects y only indirectly through x. In terms of probability, this implies the exclusion restriction of the form

$$\Pr\left[y \mid x, z\right]=\Pr\left[y \mid x\right]$$

We therefore have

- $\Pr\left[y=1 \mid x=1, z=1\right]=\Pr\left[y=1 \mid x=1, z=0\right] \quad (=\Pr\left[y=1 \mid x=1\right]) \quad =\alpha$

- $\Pr\left[y=0 \mid x=1, z=1\right]=\Pr\left[y=0 \mid x=1, z=0\right] \quad (=\Pr\left[y=0 \mid x=1\right]) \quad =1-\alpha$

- $\Pr\left[y=1 \mid x=0, z=1\right]=\Pr\left[y=1 \mid x=0, z=0\right] \quad (=\Pr\left[y=1 \mid x=0\right]) \quad =\beta$

- $\Pr\left[y=0|\, x=0,z=1\right]=\Pr\left[y=0|\, x=0,z=0\right] \quad \left(=\Pr\left[y=0|\, x=0\right]\right) \quad =1-\beta$

Using the definition of $\gamma,\delta,\varphi,\psi$, we can write the joint probabilities as

- $p(1,1,1)\equiv u_1\equiv \Pr\left[y=1,x=1,z=1\right]=\alpha\gamma$
- $p(1,1,0)\equiv u_2\equiv \Pr\left[y=1,x=1,z=0\right]=\alpha\delta$
- $p(0,1,1)\equiv u_3\equiv \Pr\left[y=0,x=1,z=1\right]=(1-\alpha)\gamma$
- $p(0,1,0)\equiv u_4\equiv \Pr\left[y=0,x=1,z=0\right]=(1-\alpha)\delta$
- $p(1,0,1)\equiv u_5\equiv \Pr\left[y=1,x=0,z=1\right]=\beta\varphi$
- $p(1,0,0)\equiv u_6\equiv \Pr\left[y=1,x=0,z=0\right]=\beta\psi$
- $p(0,0,1)\equiv u_7\equiv \Pr\left[y=0,x=0,z=1\right]=(1-\beta)\varphi$
- $p(0,0,0)\equiv u_8\equiv \Pr\left[y=0,x=0,z=0\right]=(1-\beta)\psi$

Remark 5 *We do not need to define p and u separately, but it is convenient for accounting/computer programming purpose.*

Remark 6 *Below, we will use u and $\widehat{u}$ to denote the vector of u's and corresponding sample frequencies, i.e., $u=(u_1,u_2,\ldots,u_8)'$. We will use the fact that the asymptotic variance of $\sqrt{n}\left(\widehat{u}-u\right)$ is $N(0,\Sigma)$, where*

$$\Sigma=\operatorname{diag}(u)-uu'=\left[\Sigma_1,\Sigma_2\right]$$

where

$$\Sigma_1=\begin{bmatrix} -\alpha\gamma(\alpha\gamma-1) & -\alpha^2\gamma\delta & \alpha\gamma^2(\alpha-1) & \alpha\gamma\delta(\alpha-1) \\ -\alpha^2\gamma\delta & -\alpha\delta(\alpha\delta-1) & \alpha\gamma\delta(\alpha-1) & \alpha\delta^2(\alpha-1) \\ \alpha\gamma^2(\alpha-1) & \alpha\gamma\delta(\alpha-1) & -\gamma(\alpha-1)(\alpha\gamma-\gamma+1) & -\gamma\delta(\alpha-1)^2 \\ \alpha\gamma\delta(\alpha-1) & \alpha\delta^2(\alpha-1) & -\gamma\delta(\alpha-1)^2 & -\delta(\alpha-1)(\alpha\delta-\delta+1) \\ -\alpha\beta\gamma\varphi & -\alpha\beta\delta\varphi & \beta\gamma\varphi(\alpha-1) & \beta\delta\varphi(\alpha-1) \\ -\alpha\beta\gamma\psi & -\alpha\beta\delta\psi & \beta\gamma\psi(\alpha-1) & \beta\delta\psi(\alpha-1) \\ \alpha\gamma\varphi(\beta-1) & \alpha\delta\varphi(\beta-1) & -\gamma\varphi(\alpha-1)(\beta-1) & -\delta\varphi(\alpha-1)(\beta-1) \\ \alpha\gamma\psi(\beta-1) & \alpha\delta\psi(\beta-1) & -\gamma\psi(\alpha-1)(\beta-1) & -\delta\psi(\alpha-1)(\beta-1) \end{bmatrix}$$

and

$$\Sigma_1=\begin{bmatrix} -\alpha\beta\gamma\varphi & -\alpha\beta\gamma\psi & \alpha\gamma\varphi(\beta-1) & \alpha\gamma\psi(\beta-1) \\ -\alpha\beta\delta\varphi & -\alpha\beta\delta\psi & \alpha\delta\varphi(\beta-1) & \alpha\delta\psi(\beta-1) \\ \beta\gamma\varphi(\alpha-1) & \beta\gamma\psi(\alpha-1) & -\gamma\varphi(\alpha-1)(\beta-1) & -\gamma\psi(\alpha-1)(\beta-1) \\ \beta\delta\varphi(\alpha-1) & \beta\delta\psi(\alpha-1) & -\delta\varphi(\alpha-1)(\beta-1) & -\delta\psi(\alpha-1)(\beta-1) \\ -\beta\varphi(\beta\varphi-1) & -\beta^2\psi\varphi & \beta\varphi^2(\beta-1) & \beta\psi\varphi(\beta-1) \\ -\beta^2\psi\varphi & -\beta\psi(\beta\psi-1) & \beta\psi\varphi(\beta-1) & \beta\psi^2(\beta-1) \\ \beta\varphi^2(\beta-1) & \beta\psi\varphi(\beta-1) & -\varphi(\beta-1)(\beta\varphi-\varphi+1) & -\psi\varphi(\beta-1)^2 \\ \beta\psi\varphi(\beta-1) & \beta\psi^2(\beta-1) & -\psi\varphi(\beta-1)^2 & -\psi(\beta-1)(\beta\psi-\psi+1) \end{bmatrix}$$

Natural estimators of the parameters For asymptotics, we relate the natural estimators to the sample frequencies, i.e., the u's.

$$\begin{aligned}\widehat{\alpha} &= \widehat{\Pr}\left[y=1|\, x=1\right] = \frac{\widehat{\Pr}\left[y=1, x=1\right]}{\widehat{\Pr}\left[x=1\right]} \\ &= \frac{\widehat{p}\left(1,1,1\right)+\widehat{p}\left(1,1,0\right)}{\widehat{p}\left(1,1,1\right)+\widehat{p}\left(1,1,0\right)+\widehat{p}\left(0,1,1\right)+\widehat{p}\left(0,1,0\right)} = \frac{\widehat{u}_1+\widehat{u}_2}{\widehat{u}_1+\widehat{u}_2+\widehat{u}_3+\widehat{u}_4}\end{aligned}$$

$$\begin{aligned}\widehat{\beta} &= \widehat{\Pr}\left[y=1|\, x=0\right] = \frac{\widehat{\Pr}\left[y=1, x=0\right]}{\widehat{\Pr}\left[x=0\right]} \\ &= \frac{\widehat{p}\left(1,0,1\right)+\widehat{p}\left(1,0,0\right)}{\widehat{p}\left(1,0,1\right)+\widehat{p}\left(1,0,0\right)+\widehat{p}\left(0,0,1\right)+\widehat{p}\left(0,0,0\right)} = \frac{\widehat{u}_5+\widehat{u}_6}{\widehat{u}_5+\widehat{u}_6+\widehat{u}_7+\widehat{u}_8}\end{aligned}$$

$$\begin{aligned}\widehat{\gamma} &= \widehat{\Pr}\left[x=1, z=1\right] = \widehat{p}\left(1,1,1\right)+\widehat{p}\left(0,1,1\right) \\ &= \widehat{u}_1+\widehat{u}_3\end{aligned}$$

$$\begin{aligned}\widehat{\delta} &= \widehat{\Pr}\left[x=1, z=0\right] = \widehat{p}\left(1,1,0\right)+\widehat{p}\left(0,1,0\right) \\ &= \widehat{u}_2+\widehat{u}_4\end{aligned}$$

$$\begin{aligned}\widehat{\varphi} &= \widehat{\Pr}\Pr\left[x=0, z=1\right] = \widehat{p}\left(1,0,1\right)+\widehat{p}\left(0,0,1\right) \\ &= \widehat{u}_5+\widehat{u}_7\end{aligned}$$

Understanding the two possible estimators The decomposition estimator is

$$\begin{aligned}\widehat{\theta} &= \widehat{\alpha}\widehat{\xi}+\widehat{\beta}\left(1-\widehat{\xi}\right) \\ &= \frac{\widehat{u}_1+\widehat{u}_2}{\widehat{u}_1+\widehat{u}_2+\widehat{u}_3+\widehat{u}_4}\frac{\widehat{u}_1+\widehat{u}_3}{\widehat{u}_1+\widehat{u}_3+\widehat{u}_5+\widehat{u}_7}+\frac{\widehat{u}_5+\widehat{u}_6}{\widehat{u}_5+\widehat{u}_6+\widehat{u}_7+\widehat{u}_8}\frac{\widehat{u}_5+\widehat{u}_7}{\widehat{u}_1+\widehat{u}_3+\widehat{u}_5+\widehat{u}_7}\end{aligned}$$

and the one shot estimator is

$$\begin{aligned}\widetilde{\theta} &= \widehat{\Pr}\left[y=1|\, z=1\right] = \frac{\widehat{\Pr}\left[y=1, z=1\right]}{\widehat{\Pr}\left[z=1\right]} \\ &= \frac{\widehat{\Pr}\left[y=1, x=1, z=1\right]+\widehat{\Pr}\left[y=1, x=0, z=1\right]}{\widehat{\Pr}\left[z=1\right]} \\ &= \frac{\widehat{p}\left(1,1,1\right)+\widehat{p}\left(1,0,1\right)}{\widehat{p}\left(1,1,1\right)+\widehat{p}\left(0,1,1\right)+\widehat{p}\left(1,0,1\right)+\widehat{p}\left(0,0,1\right)} \\ &= \frac{\widehat{u}_1+\widehat{u}_5}{\widehat{u}_1+\widehat{u}_3+\widehat{u}_5+\widehat{u}_7}\end{aligned}$$

Asymptotic distribution of the decomposition estimator We first establish the asymptotic distribution of $\left(\widehat{\alpha},\widehat{\beta},\widehat{\xi}\right)'$ by using delta method. We will recall that

$$\begin{aligned}\widehat{\alpha} &= \frac{\widehat{u}_1+\widehat{u}_2}{\widehat{u}_1+\widehat{u}_2+\widehat{u}_3+\widehat{u}_4}\\ \widehat{\beta} &= \frac{\widehat{u}_5+\widehat{u}_6}{\widehat{u}_5+\widehat{u}_6+\widehat{u}_7+\widehat{u}_8}\\ \widehat{\xi} &= \frac{\widehat{\gamma}}{\widehat{\gamma}+\widehat{\varphi}} = \frac{\widehat{u}_1+\widehat{u}_3}{\widehat{u}_1+\widehat{u}_3+\widehat{u}_5+\widehat{u}_7}\end{aligned}$$

We will basically understand $\left(\widehat{\alpha},\widehat{\beta},\widehat{\xi}\right)'$ as a function of $\widehat{u}$. Letting

$$\Lambda = \frac{\partial\left(\alpha,\beta,\xi\right)}{\partial u} = \begin{bmatrix} \frac{u_3+u_4}{(u_1+u_2+u_3+u_4)^2} & 0 & \frac{u_5+u_7}{(u_1+u_3+u_5+u_7)^2}\\ \frac{u_3+u_4}{(u_1+u_2+u_3+u_4)^2} & 0 & 0\\ -\frac{u_1+u_2}{(u_1+u_2+u_3+u_4)^2} & 0 & \frac{u_5+u_7}{(u_1+u_3+u_5+u_7)^2}\\ -\frac{u_1+u_2}{(u_1+u_2+u_3+u_4)^2} & 0 & 0\\ 0 & \frac{u_7+u_8}{(u_5+u_6+u_7+u_8)^2} & -\frac{u_1+u_3}{(u_1+u_3+u_5+u_7)^2}\\ 0 & \frac{u_7+u_8}{(u_5+u_6+u_7+u_8)^2} & 0\\ 0 & -\frac{u_5+u_6}{(u_5+u_6+u_7+u_8)^2} & -\frac{u_1+u_3}{(u_1+u_3+u_5+u_7)^2}\\ 0 & -\frac{u_5+u_6}{(u_5+u_6+u_7+u_8)^2} & 0 \end{bmatrix} = \begin{bmatrix} -\frac{\alpha-1}{\gamma+\delta} & 0 & \frac{\varphi}{(\gamma+\varphi)^2}\\ -\frac{\alpha-1}{\gamma+\delta} & 0 & 0\\ -\frac{\alpha}{\gamma+\delta} & 0 & \frac{\varphi}{(\gamma+\varphi)^2}\\ -\frac{\alpha}{\gamma+\delta} & 0 & 0\\ 0 & -\frac{\beta-1}{\psi+\varphi} & -\frac{\gamma}{(\gamma+\varphi)^2}\\ 0 & -\frac{\beta-1}{\psi+\varphi} & 0\\ 0 & -\frac{\beta}{\psi+\varphi} & -\frac{\gamma}{(\gamma+\varphi)^2}\\ 0 & -\frac{\beta}{\psi+\varphi} & 0 \end{bmatrix}$$

the asymptotic variance matrix is given by

$$\Lambda'\Sigma\Lambda = \begin{bmatrix} -\alpha\frac{\alpha-1}{\gamma+\delta} & 0 & 0\\ 0 & \beta\frac{\beta-1}{\gamma+\delta-1} & 0\\ 0 & 0 & \gamma\frac{\varphi}{(\gamma+\varphi)^3} \end{bmatrix}$$

We now use the delta method to

$$\widehat{\theta} = \widehat{\alpha}\widehat{\xi} + \widehat{\beta}\left(1-\widehat{\xi}\right)$$

understanding θ as a function of $\left(\widehat{\alpha},\widehat{\beta},\widehat{\xi}\right)'$. Its asymptotic variance is given by

$$\begin{aligned}&\begin{bmatrix}\xi & 1-\xi & \alpha-\beta\end{bmatrix}\begin{bmatrix} -\alpha\frac{\alpha-1}{\gamma+\delta} & 0 & 0\\ 0 & \beta\frac{\beta-1}{\gamma+\delta-1} & 0\\ 0 & 0 & \gamma\frac{\varphi}{(\gamma+\varphi)^3} \end{bmatrix}\begin{bmatrix}\xi\\ 1-\xi\\ \alpha-\beta\end{bmatrix}\\ &\qquad = \frac{\alpha\left(1-\alpha\right)\xi^2}{\gamma+\delta} + \frac{\beta\left(1-\beta\right)\left(1-\xi\right)^2}{1-\gamma-\delta} + \frac{\gamma\varphi\left(\alpha-\beta\right)^2}{\left(\gamma+\varphi\right)^3}\end{aligned}$$

In other words, the asymptotic variance of the compositional estimator is equal to

$$\frac{\alpha\left(1-\alpha\right)\xi^2}{\gamma+\delta} + \frac{\beta\left(1-\beta\right)\left(1-\xi\right)^2}{1-\gamma-\delta} + \frac{\gamma\varphi\left(\alpha-\beta\right)^2}{\left(\gamma+\varphi\right)^3} \tag{16}$$

Asymptotic distribution of the one shot estimator We will basically understand

$$\widetilde{\theta} = \frac{\widehat{u}_1 + \widehat{u}_5}{\widehat{u}_1 + \widehat{u}_3 + \widehat{u}_5 + \widehat{u}_7}$$

as a function of $\widehat{u}$. Letting

$$\frac{\partial}{\partial u_1}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = \frac{u_3 + u_7}{\left(u_1 + u_3 + u_5 + u_7\right)^2}$$

$$\frac{\partial}{\partial u_2}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = 0$$

$$\frac{\partial}{\partial u_3}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = -\frac{u_1 + u_5}{\left(u_1 + u_3 + u_5 + u_7\right)^2}$$

$$\frac{\partial}{\partial u_4}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = 0$$

$$\frac{\partial}{\partial u_5}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = \frac{u_3 + u_7}{\left(u_1 + u_3 + u_5 + u_7\right)^2}$$

$$\frac{\partial}{\partial u_6}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = 0$$

$$\frac{\partial}{\partial u_7}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = -\frac{u_1 + u_5}{\left(u_1 + u_3 + u_5 + u_7\right)^2}$$

$$\frac{\partial}{\partial u_8}\left(\frac{u_1 + u_5}{u_1 + u_3 + u_5 + u_7}\right) = 0$$

In order to apply the delta method, we stack these expressions as an eight dimensional column vector, and evaluate at the true value of the u's, e.g., $u_1 = \alpha\gamma\varphi$. We then obtain a column vector

$$\chi = \frac{\partial\left(\frac{u_1+u_5}{u_1+u_3+u_5+u_7}\right)}{\partial u} = \begin{bmatrix} \frac{u_3+u_7}{(u_1+u_3+u_5+u_7)^2} \\ 0 \\ -\frac{u_1+u_5}{(u_1+u_3+u_5+u_7)^2} \\ 0 \\ \frac{u_3+u_7}{(u_1+u_3+u_5+u_7)^2} \\ 0 \\ -\frac{u_1+u_5}{(u_1+u_3+u_5+u_7)^2} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{(\gamma+\varphi)^2}\left(\gamma + \varphi - \alpha\gamma - \beta\varphi\right) \\ 0 \\ -\frac{1}{(\gamma+\varphi)^2}\left(\alpha\gamma + \beta\varphi\right) \\ 0 \\ \frac{1}{(\gamma+\varphi)^2}\left(\gamma + \varphi - \alpha\gamma - \beta\varphi\right) \\ 0 \\ -\frac{1}{(\gamma+\varphi)^2}\left(\alpha\gamma + \beta\varphi\right) \\ 0 \end{bmatrix}$$

Using the delta method, we can conclude that the asymptotic variance of the one shot estimator is equal to

$$\chi'\Sigma\chi = \frac{1}{\left(\gamma + \varphi\right)^3}\left(\alpha\gamma + \beta\varphi\right)\left(\gamma + \varphi - \alpha\gamma - \beta\varphi\right)$$

In other words, the asymptotic variance of the one shot estimator is equal to

$$\frac{1}{(\gamma+\varphi)^3}(\alpha\gamma+\beta\varphi)(\gamma+\varphi-\alpha\gamma-\beta\varphi) \tag{17}$$

Subtracting (16) from (17), we obtain

$$\begin{aligned}(17)-(16 &= \frac{\varphi(\gamma+\delta)(1-(\gamma+\delta+\varphi))\beta(1-\beta)+\gamma\delta(1-(\gamma+\delta))\alpha(1-\alpha)}{(\gamma+\varphi)^2(1-(\gamma+\delta))(\gamma+\delta)} \\ &= \frac{\varphi(\gamma+\delta)\psi\beta(1-\beta)+\gamma\delta(\varphi+\psi)\alpha(1-\alpha)}{(\gamma+\varphi)^2(\varphi+\psi)(\gamma+\delta)} > 0\end{aligned}$$

In other words, the compositional estimator is more efficient than the one shot estimator.