Omitted Variables, Instrumental Variables (IV), and Two-Stage Least Squares (TSLS)

Greene Ch.8, 12, Kennedy Ch. 9
R script `mod4s1a, mod4s1b, mod4s1c`

Assumption 3 of the CLRM stipulates that the explanatory variables are uncorrelated with the error term. In many real-world applications this assumption will not hold. Examples include:

- Omitted variable (a variable that affects **y** AND is correlated with one or more regressors is omitted from the model).
- Measurement error on one or more regressors
- Lagged dependent variables used as regressors ("autoregression")
- Simultaneous equations
- Models with sample selection

All these conditions will result in the same econometric problem: biasedness and inconsistency of the OLS (or MLE) estimator. For this chapter we will focus on least squares (LS) estimation and the remedial methods of IV and TSLS that have been devised for the LS setting.

Orthogonality and Covariance

In the following, we will make ample use of the concepts of "orthogonality" , sample covariance, and sample variance. Let's take a closer look at these terms.

Consider two nx1 vectors **x** and **z**. Orthogonality implies that

$$\mathbf{x} \perp \mathbf{z} \rightarrow \mathbf{x}'\mathbf{z} = \sum_{i=1}^{n} x_i z_i = 0 \tag{1}$$

The sample variance of x is given as

$$\begin{aligned} \operatorname{var}(\mathbf{x}) = s_x^2 = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = (n-1)^{-1} \left(\sum_{i=1}^{n} x_i^2 - n^{-1} \left(\sum_{i=1}^{n} x_i \right)^2 \right) = \\ (n(n-1))^{-1} \left(n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right) \end{aligned} \tag{2}$$

The last expression will be useful below. The sample covariance between **x** and **z** can be written as:

$$\operatorname{cov}(\mathbf{x},\mathbf{z}) = s_{xz} = (n-1)^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z}) = (n-1)^{-1}\left(\sum_{i=1}^{n}x_i z_i - n\overline{xz}\right) =$$
$$(n(n-1))^{-1}\left(n\sum_{i=1}^{n}x_i z_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}z_i\right) \quad (3)$$

Again, the last expression will be useful later.

It follows that orthogonality alone does NOT imply a zero covariance. For a zero covariance we need either $n\sum_{i=1}^{n}x_i z_i = \sum_{i=1}^{n}x_i\sum_{i=1}^{n}z_i$ (a somewhat unlikely condition), or either of the sums in the last term is zero in addition to orthogonality. By the same token, strictly speaking a zero sample covariance does not imply orthogonality. However, orthogonality likely holds in that case, otherwise we would need again $n\sum_{i=1}^{n}x_i z_i = \sum_{i=1}^{n}x_i\sum_{i=1}^{n}z_i$ for the covariance to be zero.

We should also note that the sample variance converges to the population variance (call it σ_x^2) as *n* goes to infinity. We can show this via convergence in mean square error:

$$E\left(s_x^2\right) = (n-1)^{-1}\left(E\left(\sum_{i=1}^{n}x_i^2\right) - nE\left(\bar{x}^2\right)\right) = n(n-1)^{-1}\left(E\left(x_i^2\right) - E\left(\bar{x}^2\right)\right) =$$
$$n(n-1)^{-1}\left(V(x_i) + \left(E(x_i)\right)^2 - V(\bar{x}) - E(\bar{x})^2\right) = \quad (4)$$
$$n(n-1)^{-1}\left(\sigma_x^2 + \mu_x^2 - \tfrac{1}{n}\sigma_x^2 - \mu_x^2\right) = \frac{n(n-1)}{n(n-1)}\sigma_x^2 = \sigma_x^2$$

Thus $\lim_{n\to\infty} E\left(s_x^2\right) = \sigma_x^2$. Similarly, it can be shown that $\lim_{n\to\infty} \operatorname{var}\left(s_x^2\right) = 0$. Thus $\operatorname{plim} s_x^2 = \sigma_x^2$. See also Greene p. 67-68. By analogy, the sample covariance approaches the population covariance (call it σ_{xz}) under increasing sample size.

In the following, when we examine the *covariance* (or correlation) between included regressors and the components of the error term we usually think of this as the population covariance. However, some concepts are more clearly illustrated using the sample covariance. Either one is suitable in illustrating the O.V. problem.

The Omitted variable problem

Consider the stipulated model $\mathbf{y} = \mathbf{X_1\beta_1} + \boldsymbol{\varepsilon}$ and the true model $\mathbf{y} = \mathbf{X_1\beta_1} + \mathbf{X_2\beta_2} + \boldsymbol{\nu}$. Asssume for now that $E\left(\mathbf{X_1}'\boldsymbol{\nu}\right) = E\left(\mathbf{X_2}'\boldsymbol{\nu}\right) = \mathbf{0}$. The first model, estimated via OLS, produces

$$\mathbf{b_1} = \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'y} = \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'}\left(\mathbf{X_1\beta_1} + \mathbf{X_2\beta_2} + \mathbf{v}\right) =$$
$$\boldsymbol{\beta_1} + \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'X_2\beta_2} + \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'v} \quad \text{with} \tag{5}$$
$$\mathrm{E}\left(\mathbf{b_1} \mid \mathbf{X}\right) = \boldsymbol{\beta_1} + \mathbf{P_{1.2}\beta_2} \quad \text{where} \quad \mathbf{P_{1.2}} = \left(\mathbf{X_1'X_1}\right)^{-1}\mathbf{X_1'X_2}$$

The last term in (5) will only be zero if the two **X**-matrices are perfectly orthogonal *or* if $\boldsymbol{\beta_2} = \mathbf{0}$, i.e. the omitted variables have no effect on **y (**which, of course, we ruled out upfront when we wrote down the "true" model). Alternatively you can think of each column in $\mathbf{P_{1.2}}$ as a set of LS coefficients from a regression of the corresponding column of $\mathbf{X_2}$ on the entire $\mathbf{X_1}$ matrix. In other word, the stipulated or "flawed model" will lead to unbiased (and consistent) estimates only if the included regressors ($\mathbf{X_1}$) are orthogonal (and thus, in essence, uncorrelated) with the excluded variables ($\mathbf{X_2}$) and thus with the error term $\boldsymbol{\varepsilon} = \mathbf{X_2\beta_2} + \mathbf{v}$.

Here is a closer look at this situation using a simple stipulated model with a constant term and a single regressor, and a single omitted variable. Thus, we have

$$\begin{aligned} &\mathbf{y} = \beta_1\mathbf{i} + \beta_2\mathbf{x} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} = \gamma\mathbf{z} + \nu \qquad \text{or} \\ &\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \mathbf{X} = \left[\mathbf{i} \quad \mathbf{x}\right] \end{aligned} \tag{6}$$

This implies

$$\mathbf{P_{X.Z}} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'z} = \begin{bmatrix} \dfrac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} & \dfrac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ \dfrac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} & \dfrac{n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n z_i \\ \sum_{i=1}^n x_i z_i \end{bmatrix} =$$

$$\begin{bmatrix} \dfrac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n z_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i z_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ \dfrac{n\sum_{i=1}^n x_i z_i - \sum_{i=1}^n x_i \sum_{i=1}^n z_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \end{bmatrix} = \begin{bmatrix} \dfrac{(n-1)^{-1}\left(\sum_{i=1}^n x_i \bar{z} - \sum_{i=1}^n x_i z_i \bar{x}\right)}{s_x^2} \\ \dfrac{s_{xz}}{s_x^2} \end{bmatrix} \tag{7}$$

where we use the results from (2) and (3) in the last transformation. This shows some additional details for the omitted variable problem. If the sample covariance between **x** and **z** is zero, only the intercept will be biased. If **x** and **z** are orthogonal (or have zero covariance) and $\sum_{i=1}^{n} z_i = 0$, neither term will be biased. Thus, the omitted variable problem can be quite subtle even for this simple case, assuming $\gamma \neq 0$.

In a multivariate regression model, O.V. bias will generally affect ALL coefficients in unknown direction, even if there is only one O.V. correlated with only one regressor! This is due to the usual correlation between included regressors (i.e. $\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}$ will usually be a full *k* by *k* matrix with nonzero elements) .

Returning to the general case, we can alternatively express this problem as a violation of Assumption 3 of the CLRM, i.e.

$$E\left(\mathbf{X}'\boldsymbol{\varepsilon}\right) = \boldsymbol{\gamma} \neq \mathbf{0} \tag{8}$$

As a result, the OLS estimator is biased and inconsistent (i.e. asymptotically biased), i.e.

$$\begin{aligned} &E\left(\mathbf{b} \mid \mathbf{X}\right) = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\boldsymbol{\gamma} \\ &\text{plim}\,\mathbf{b} = \boldsymbol{\beta} + \text{plim}\left(\tfrac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\text{plim}\left(\tfrac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon}\right) = \boldsymbol{\beta} + \mathbf{Q}_{\mathbf{XX}}^{-1}\boldsymbol{\gamma} \end{aligned} \tag{9}$$

R script `mod4s1a` provides an empirical illustration of the O.V. problem using the hedonic property value data from PS3, Q3.

The Instrumental Variables (IV) Estimator and TSLS

Assume there are k_1 columns of (*n* by *k*) matrix **X** that are potentially correlated with the error term. We will call these variables "*troublemakers*". Assume the remaining k_2 columns are "clean" with respect to O.V. problems (although not necessarily uncorrelated with the remaining columns in **X**). Assume that for each troublemaker column $\mathbf{c_j}$ we can find *one or more variables* that are not in the original models, and that are highly correlated with $\mathbf{c_j}$, but not with $\boldsymbol{\varepsilon}$. Such variables are called "*instruments*". Instruments are "crutches" that ultimately allow us to derive the effect of the original **X** on **y** without incurring O.V. bias. Thus, clean, or "valid" instruments are not correlated with the outcome variable **y** other than through their correlation with $\mathbf{c_j}$.

Let matrix **Z** collect all "clean" columns of the original **X**, *plus* the instruments. Assume **Z** has dimensions n by *l*, with $l \geq k$.

Next, consider the following *2-step procedure*: First, the troublemakers in **X** need to be purged of their O.V.-causing effects. Second, the purged **X** is used in the originally desired regression.

The first step is accomplished by regressing each column in **X** (troublemaker or not) against **Z** and compute fitted values. This will leave the clean columns of **X** unchanged, but leaves the troublemakers with only the effects that can *also* be explained by **Z**. Formally:

$$\hat{\mathbf{X}} = \mathbf{P_Z X} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X} \tag{10}$$

In the second step, $\hat{\mathbf{X}}$ is used in lieu of $\mathbf{X}$ in the original regression, and OLS is applied. This leads to the Two-Stage-Least-Squares (TSLS) Estimator

$$\mathbf{b_{TSLS}} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y} = \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y} \tag{11}$$

This expression can be simplified in the special case where $l = k$, i.e. each troublemaker is replaced by exactly *one* corresponding instrument. In that case, $\mathbf{Z}'\mathbf{X}$ is a square matrix with a (presumably) well-behaved inverse, and we can write

$$\mathbf{b_{TSLS}} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{Z}'\mathbf{Z}\right)\left(\mathbf{X}'\mathbf{Z}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{y} = \mathbf{b_{IV}} \tag{12}$$

In this special case the TSLS estimator is usually called "IV estimator" or $\mathbf{b_{IV}}$. However, the two terms are often used interchangeably. Just remember that the expression in (11) always holds, while (12) only holds if $l = k$.

As shown in Greene, Ch. 8, $\mathbf{b_{IV}}$ (used in the general sense of $\mathbf{b_{TSLS}}$) is consistent and asymptotically normally distributed. It's asymptotic variance can be estimated by

$$\hat{V}_a\left(\mathbf{b_{TSLS}}\right) = \hat{\sigma}^2\left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1} = \hat{\sigma}^2\left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1} \qquad \text{where} \quad \hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n}$$
$$\text{and} \quad \hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\mathbf{b_{TSLS}} \tag{13}$$

Note that the *original* $\mathbf{X}$, not $\hat{\mathbf{X}}$, is used in the computation of the residuals $\hat{\boldsymbol{\varepsilon}}$!

The asymptotic variance of the TSLS estimator can shown to be "larger" than that of the OLS estimator, especially when the instruments are "poor" (i.e. not highly correlated with the troublemaker(s)). In other words, the TSLS estimator is less efficient than the OLS estimator. However, efficiency is not a very meaningful virtue if the estimator is inconsistent!

Specification tests for OV-type problems

Hausman Test (H-test in my jargon)

The Hausman test is a type of Wald test that examines if the difference between 2 sets of estimates flowing from 2 different models, weighted by the difference in their asymptotic variance-covariance matrix, is "large enough" to reject the null hypothesis that they are the same.

The rationale is that the first model considered is known to generate *consistent* estimates under OV-type problems or other mis-specification issues, while the second model is *inconsistent* IF there are indeed OV type problems or mis-specifications. However, the first estimator is always *less efficient* than the second. So if there are *no* OV-type or mis-specification problems, it would be better to choose the second model. If there are OV type problems, we should use the first model (since consistency is generally more important than efficiency).

For the case at hand the IV (or TSLS) model is "model 1" – consistent under OV-problems, but less efficient. The OLS model is "model 2" – more efficient, but inconsistent under OV-problems. The H-test examines if the two estimators are "close enough" to conclude that OLS is fine, i.e. that there are no OV-type problems (this is the null hypothesis). If the weighted difference between the estimators is "too large", the test would reject the null.

The H-test statistic is thus derived as

$$H = \mathbf{d}'\left(\hat{V}_a\left(\mathbf{b}_{\mathbf{TSLS}}\right) - \hat{V}_a\left(\mathbf{b}\right)\right)^{-1}\mathbf{d} = \mathbf{d}'\left(s^2\left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1} - s^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}\mathbf{d} \sim \chi^2\left(J\right)$$
$$\text{where} \quad \mathbf{d} = \left(\mathbf{b}_{\mathbf{TSLS}} - \mathbf{b}\right), \quad \hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}, \quad s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} \tag{14}$$

By convention the squared regression error (s^2) from the OLS model is used for both variances in this formula. Being a type of Wald test, the test statistic is distributed chi-square with J degrees of freedom, where J is equal to the *number of "troublemakers*" in the original model (i.e. the columns of $\mathbf{X}$ that are being replaced by instruments).

A common problem that arises with this test is that $\left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right) - \left(\mathbf{X}'\mathbf{X}\right)$ may not be full rank since there are usually some common columns in $\hat{\mathbf{X}}$ and $\mathbf{X}$**.** Thus, the inverse of the difference in variances may not exist. To get around this problem the "*Moore-Penrose Generalized Inverse*" is used instead. In **R**, just use `pseudoinverse` (in the `corpcor` library) instead of `solve`.

Wu test

An equivalent test that avoids this "inverse" problem is the Wu test. It proceeds in 2 steps:

1. Pick the troublemakers from $\mathbf{X}$ and collect them in a new matrix, say $\mathbf{X}^*$. Regress each column in $\mathbf{X}^*$ against $\mathbf{Z}$ and obtain the fitted values, i.e. generate $\hat{\mathbf{X}}^* = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}^*$.
2. Regress $\mathbf{y}$ against the original $\mathbf{X}$ and $\hat{\mathbf{X}}^*$.

The null hypothesis for the test is that the coefficients for $\hat{\mathbf{X}}^*$ are jointly zero. A rejection of the null would indicate OV problems in the OLS model, and would suggest switching to an IV approach. If the null is not rejected, OLS is fine.

This is implemented as an F-test with (n-k) and J degrees of freedom, where J is again the number of troublemakers, here equal to the columns in $\mathbf{X}^*$ and k is the number of regressors in the final model in step 2.

Examples

Example 1: OV-type problems through endogeneity

Endogeneity bias arises when one or more explanantory variables in the main equation of interest are themselves directly or indirectly driven by the dependent variable. This is also known as "simultaneous equation" bias. Often times the variables are linked by some underlying theoretical relationship (in which case we can usually anticipate trouble...). The upshot is that this endogeneity introduces a correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, rendering the OLS estimator biased and inconsistent.

Script `mod4s1b` illustrates this effect. It is based on Greene's (5th edition) example 5.3.

The main equation of interest is a regression of national consumption expenditures at time *t* against national disposable income at time *t*, i.e.

$$y_t = \beta_1 + \beta_2 dpi_t + \varepsilon_t \tag{15}$$

However, the national accounting identity stipulates that national income is the sum of consumption, investment, government spending, and net exports. Thus, at least at the macroeconomic level, income may well be driven by consumption. Let's assume for simplicity that this relationship is again linear, such that:

$$dpi_t = \gamma_1 + \gamma_2 y_t + \mathbf{x}_t'\boldsymbol{\delta} + \mu_t \tag{16}$$

where $\mathbf{x}_t$ includes other variables that affect dpi_t, and μ_t is a well behaved CLRM error term. Plugging the first equation into the second, we get

$$dpi_t = \gamma_1 + \gamma_2\left(\beta_1 + \beta_2 dpi_t + \varepsilon_t\right) + \mathbf{x}_t'\boldsymbol{\delta} + \mu_t \rightarrow dpi_t = \left(\frac{\gamma_1 + \gamma_2\beta_1}{1-\gamma_2\beta_2}\right) + \left(\frac{\mathbf{x}_t'\boldsymbol{\delta} + \mu_t}{1-\gamma_2\beta_2}\right) + \left(\frac{\gamma_2}{1-\gamma_2\beta_2}\varepsilon_t\right) \tag{17}$$

For the first equation, this implies that

$$COV\left(dpi_t, \varepsilon_t\right) = COV\left(\left(\frac{\gamma_1 + \gamma_2\beta_0}{1-\gamma_2\beta_2}\right) + \left(\frac{\mathbf{x}_t'\boldsymbol{\delta} + \mu_t}{1-\gamma_2\beta_2}\right) + \left(\frac{\gamma_2}{1-\gamma_2\beta_2}\varepsilon_t\right), \varepsilon_t\right) = \frac{\gamma_2}{1-\gamma_2\beta_2}\sigma_\varepsilon^2 \neq 0 \tag{18}$$

As can be seen from the last term the strength and direction of the link between dpi_t, and the error term depends on the magnitude and sign of γ_2 and β_2.

Example 2: OV-type problems through measurement error

Assume the correct form of your CLRM equation of interest is given by

$$y_i = \beta_1 + \beta_2 x_{2,i}^* + \mathbf{x}_{\mathbf{3:k,i}}'\boldsymbol{\beta}_{\mathbf{3:k}} + \varepsilon_i, \tag{19}$$

Further assume that regressor $\mathbf{x}_2^*$ is measured with error for the entire sample (or at least for more than a handful of observations). We only observe $\mathbf{x}_2$. Assume that the relationship between $\mathbf{x}_2$ and $\mathbf{x}_2^*$ is given as

$$\mathbf{x}_2 = \mathbf{x}_2^* + \boldsymbol{\mu} \quad with\ \boldsymbol{\mu} \sim n\left(\mathbf{0}, \sigma_\mu^2\mathbf{I}\right) \tag{20}$$

This implies that the measurement error is perfectly random. Other types of measurement error are possible. Either way, we run into the following dilemma (shown for simplicity at the individual level):

$$
\begin{aligned}
&y_i = \beta_1 + \beta_2\left(x_{2i} - \mu_i\right) + \mathbf{x}'_{\mathbf{3:k,i}}\boldsymbol{\beta}_{\mathbf{3:k}} + \varepsilon_i = \beta_1 + \beta_2 x_{2i} + \mathbf{x}'_{\mathbf{3:k,i}}\boldsymbol{\beta}_{\mathbf{3:k}} + \left(\varepsilon_i - \beta_2\mu_i\right) = \\
&\beta_1 + \beta_2 x_{2i} + \mathbf{x}'_{\mathbf{3:k,i}}\boldsymbol{\beta}_{\mathbf{3:k}} + \omega_i \qquad \text{and} \\
&\mathrm{COV}\left(x_{2i}, \omega_i\right) = COV\left(x^*_{2i} + \mu_i, \varepsilon_i - \beta_2\mu_i\right) = -\beta_2\sigma_u^2
\end{aligned} \tag{21}
$$

Again, we end up with a nonzero correlation between an included regressor and the error term of the estimated regression.

Similar problems arise if multiple columns of **X** are measured with error. As for the "classic OV case", a single mis-measured variable generally biases all estimated coefficients (unless the mis-measured variable is perfectly orthogonal with the rest of **X**).

The remedy is again to "cleanse" each "troublemaker" with one or more instruments. In this case, a good instrument will be correlated with the troublemaker, but not with either the regression error ($\boldsymbol{\varepsilon}$) or the measurement error ($\boldsymbol{\mu}$).

R script `mod4s1c` provides an example.