

# Supplementary Material (updated 21/07/20)

## Contents

A1: Derivation of the MR-GxE Estimator (equation 8).....	2
A2: Derivation of the violations of the constant pleiotropy assumption through differing confounding structures.....	3
A3: Code for performing analyses (using a preformatted data frame). .....	11

## A1: Derivation of the MR-GxE Estimator (equation 8)

Data Generating Model

$$G_i \sim \text{Normal} \quad (1)$$

$$U_i \sim \text{Normal} \quad (2)$$

$$Z_i \sim \text{Normal} \quad (3)$$

$$G_i Z_i = G_i \times Z_i \quad (4)$$

$$X_i = \gamma_0 + \gamma_1 G_i + \gamma_2 Z_i + \gamma_3 G_i Z_i + \gamma_4 U_i + \epsilon_{Xi} \quad (5)$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 G_i + \beta_3 Z_i + \beta_4 G_i Z_i + \beta_5 U_i + \epsilon_{Yi} \quad (6)$$

The MR-GxE Estimator

$$Y_i = \hat{\beta}_0 + \hat{\Gamma}_1 G_i + \hat{\Gamma}_2 Z_i + \hat{\Gamma}_3 G_i Z_i + \hat{\epsilon}_{Yi} \quad (7)$$

$$Y_i = \tilde{\Gamma}_0 + \tilde{\Gamma}_{GZ} G_i Z_i + \tilde{\epsilon}_{Yi} \quad (8)$$

$$X_i = \hat{\gamma}_0 + \hat{\gamma}_1 G_i + \hat{\gamma}_2 Z_i + \hat{\gamma}_3 G_i Z_i + \hat{\epsilon}_{Xi} \quad (9)$$

$$X_i = \tilde{\gamma}_0 + \tilde{\gamma}_{GZ} G_i Z_i + \tilde{\epsilon}_{Xi} \quad (10)$$

$$G_i = \tilde{\alpha}_0 + \tilde{\alpha}_G G_i Z_i + \tilde{\eta}_{Gi} \quad (11)$$

$$Z_i = \tilde{\alpha}_0 + \tilde{\alpha}_Z G_i Z_i + \tilde{\eta}_{Zi} \quad (12)$$

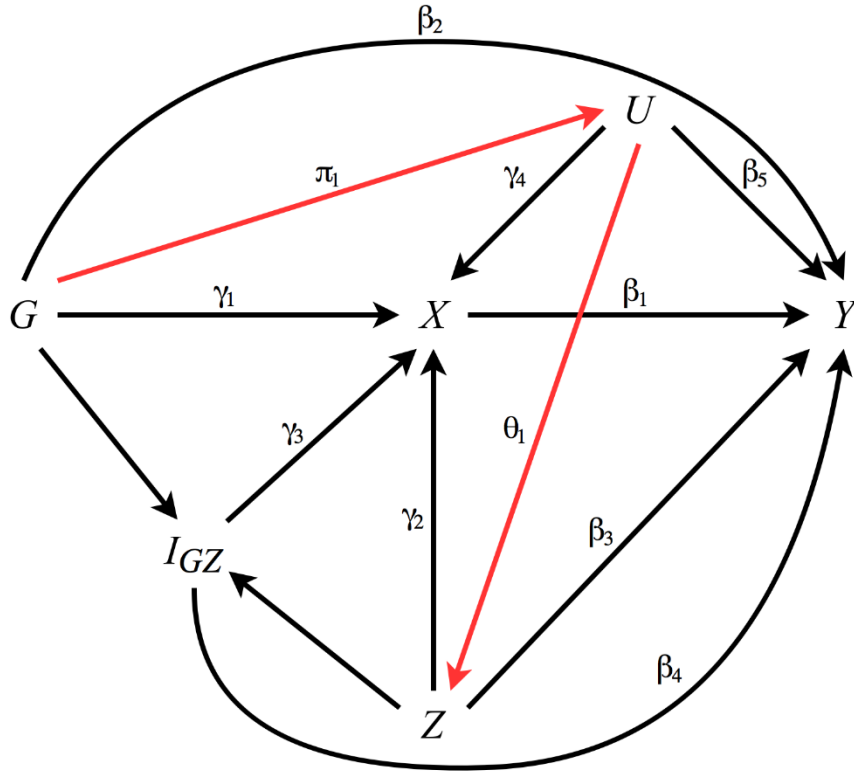
$$\hat{\beta}_{1MRGxE} = \frac{\tilde{\Gamma}_{GZ} - \tilde{\Gamma}_1 \tilde{\alpha}_G - \tilde{\Gamma}_2 \tilde{\alpha}_Z}{\tilde{\gamma}_{GZ} - \tilde{\gamma}_1 \tilde{\alpha}_G - \tilde{\gamma}_2 \tilde{\alpha}_Z} \quad (13)$$

$$\hat{\beta}_{1MRGxE} = \frac{\frac{\text{cov}(Y,GZ) - \tilde{\Gamma}_1 \text{cov}(G,GZ) - \tilde{\Gamma}_2 \text{cov}(Z,GZ)}{\text{var}(GZ)}}{\frac{\text{cov}(X,GZ) - \tilde{\gamma}_1 \text{cov}(G,GZ) - \tilde{\gamma}_2 \text{cov}(Z,GZ)}{\text{var}(GZ)}} \quad (14)$$

$$\hat{\beta}_{1MRGxE} = \frac{\text{cov}(Y,GZ) - \tilde{\Gamma}_1 \text{cov}(G,GZ) - \tilde{\Gamma}_2 \text{cov}(Z,GZ)}{\text{cov}(X,GZ) - \tilde{\gamma}_1 \text{cov}(G,GZ) - \tilde{\gamma}_2 \text{cov}(Z,GZ)} \quad (15)$$

A2: Derivation of the violations of the constant pleiotropy assumption through differing confounding structures.

**Scenario 1**



**Figure A5:** DAG illustrating a case in which the instrument  $G$  is a determinant of the interaction-covariate  $Z$  through a confounder  $U$ .

Data generating model

$$G_i = N(0,1) \tag{A21-1}$$

$$U_i = \pi_0 + \pi_1 G_i + \epsilon_U \tag{A21-2}$$

$$Z_i = \theta_0 + \theta_1 U_i + \epsilon_Z \tag{A21-3}$$

$$X_i = \gamma_0 + \gamma_1 G_i + \gamma_2 Z_i + \gamma_3 GZ_i + \gamma_4 U + \epsilon_X \tag{A21-4}$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 G_i + \beta_3 Z_i + \beta_4 GZ_i + \beta_5 U_i + \epsilon_Y \tag{A21-5}$$

Definition of bias term

First a model for  $X$  in terms of  $G$  is constructed by substituting models (A21-2) and (A21-3) into model (A21-4):

$$X_i = G_i(\gamma_1 + \gamma_2\theta_1\pi_1 + \gamma_3\theta_1\pi_1G_i + \gamma_4\pi_1) + \eta_X \tag{A21-6}$$

We can also construct a model for the outcome  $Y$  in terms of  $G$  by substituting models (A21-2), (A21-3), and (A21-6) into model (A21-5):

$$Y_i = G_i(\beta_1(\gamma_1 + \gamma_2\theta_1\pi_1 + \gamma_3\theta_1\pi_1G_i + \gamma_4\pi_1) + \beta_2 + \beta_3\theta_1\pi_1 + \beta_4\theta_1\pi_1G_i + \beta_4\pi_4) + \eta_Y \quad (\text{A21-7})$$

Using models (A21-6) and (A21-7), we can calculate the partial effect of  $G$  for the first and second stage models respectively:

$$\frac{dX}{dG} = \gamma_1 + \gamma_2\theta_1\pi_1 + 2\gamma_3\theta_1\pi_1G_i + \gamma_4\pi_1 \quad (\text{A21-8})$$

$$\frac{dY}{dG} = \beta_1(\gamma_1 + \gamma_2\theta_1\pi_1 + 2\gamma_3\theta_1\pi_1G_i + \gamma_4\pi_1) + \beta_2 + \beta_3\theta_1\pi_1 + 2\beta_4\theta_1\pi_1G_i + \beta_5\pi_1 \quad (\text{A21-9})$$

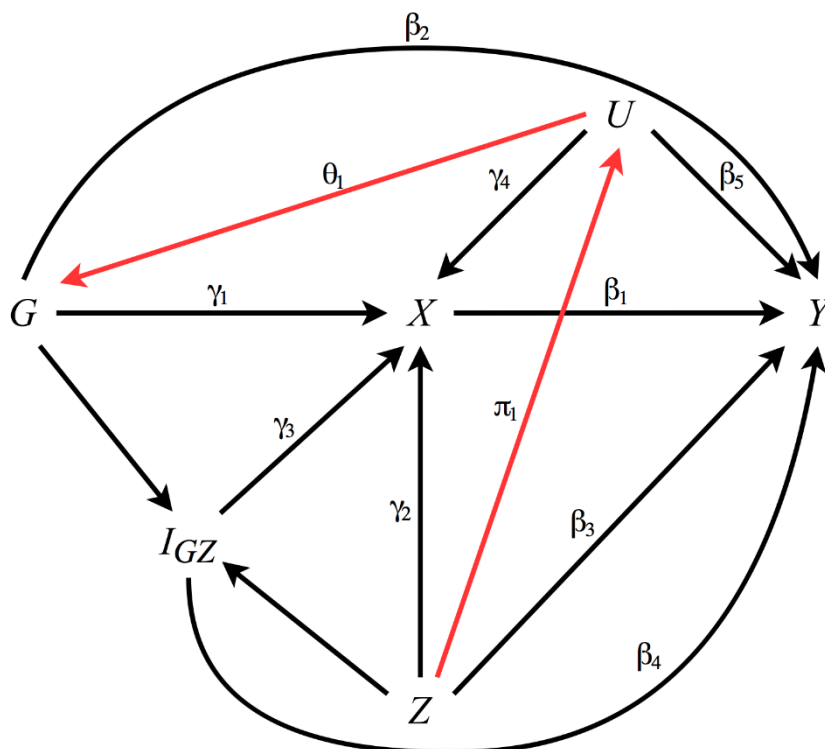
The corresponding Wald estimand is then given as:

$$\beta_1 + \frac{\beta_2 + \beta_3\theta_1\pi_1 + 2\beta_4\theta_1\pi_1G_i + \beta_5\pi_1}{\gamma_1 + \gamma_2\theta_1\pi_1 + 2\gamma_3\theta_1\pi_1G_i + \gamma_4\pi_1} \quad (\text{A21-10})$$

In equation (A21-10), the change in pleiotropic effect across the set of subgroups is represented by the term  $2\beta_4\theta_1\pi_1G_i$ . In this case,  $\beta_4$  represents the average change in pleiotropic effect across subgroups,  $\theta_1$  is the effect of  $G$  upon the confounder  $U$ , and  $\pi_1$  is the effect of the  $G$  mediated by the confounder  $U$ . This highlights three important elements of the constant pleiotropy assumption:

1. MRGXE will give an unbiased effect estimate  $\beta_1$  in cases where there is no average change in pleiotropic effect across the set of subgroups ( $\beta_4 = 0$ ). Importantly, this includes the possibility of changes in pleiotropic effect being balanced across the sample, in a similar fashion to the balanced pleiotropic effects in the MR Egger framework.
2. MRGXE will give an unbiased effect estimate  $\beta_1$  when there is no effect of the instrument  $G$  upon a confounder ( $\theta_1 = 0$ ).
3. MRGXE will give an unbiased effect estimate  $\beta_1$  when there is no effect of the instrument  $G$  mediated by the confounder ( $\pi_1 = 0$ ).

**Scenario 2:**



**Figure A6:** DAG illustrating a case in which the interaction-covariate  $Z$  is a determinant of the instrument  $G$  through a confounder  $U$ .

Data generating model

$$Z_i = N(0,1) \quad (\text{A22-1})$$

$$U_i = \pi_0 + \pi_1 Z_i + \epsilon_U \quad (\text{A22-2})$$

$$G_i = \theta_0 + \theta_1 U_i + \epsilon_G \quad (\text{A22-3})$$

$$X_i = \gamma_0 + \gamma_1 G_i + \gamma_2 Z_i + \gamma_3 GZ_i + \gamma_4 U + \epsilon_X \quad (\text{A22-4})$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 G_i + \beta_3 Z_i + \beta_4 GZ_i + \beta_5 U_i + \epsilon_Y \quad (\text{A22-5})$$

Definition of bias term

First model for the exposure  $X$  in terms of  $Z$  is constructed by substituting models (A22-2) and (A22-3) into model (A22-4):

$$X_i = Z_i(\gamma_1 \theta_1 \pi_1 + \gamma_2 + \gamma_3 \theta_1 \pi_1 Z_i + \gamma_4 \pi_1) + \eta_X \quad (\text{A22-6})$$

We can also construct a model for the outcome  $Y$  in terms of  $Z_i$  by substituting models (A22-2), (A22-3), and (A22-6) into model (A22-5):

$$Y_i = Z_i(\beta_1(\gamma_1 \theta_1 \pi_1 + \gamma_2 + \gamma_3 \theta_1 \pi_1 Z_i + \gamma_4 \pi_1) + \beta_2 \theta_1 \pi_1 + \beta_3 + \beta_4 \theta_1 \pi_1 Z_i + \beta_5 \pi_1) + \eta_Y \quad (\text{A22-7})$$

As  $Z$  is a determinant of both  $U$  and  $G$ , it is appropriate to calculate the partial effect of  $Z$  with respect to  $X$  and  $Y$

$$\frac{dX}{dZ} = \gamma_1\theta_1\pi_1 + \gamma_2 + 2\gamma_3\theta_1\pi_1Z_i + \gamma_4\pi_1 \quad (\text{A22-8})$$

$$\frac{dY}{dZ} = \beta_1(\gamma_1\theta_1\pi_1 + \gamma_2 + 2\gamma_3\theta_1\pi_1Z_i + \gamma_4\pi_1) + \beta_2\theta_1\pi_1 + \beta_3 + 2\beta_4\theta_1\pi_1Z_i + \beta_5\pi_1 \quad (\text{A22-9})$$

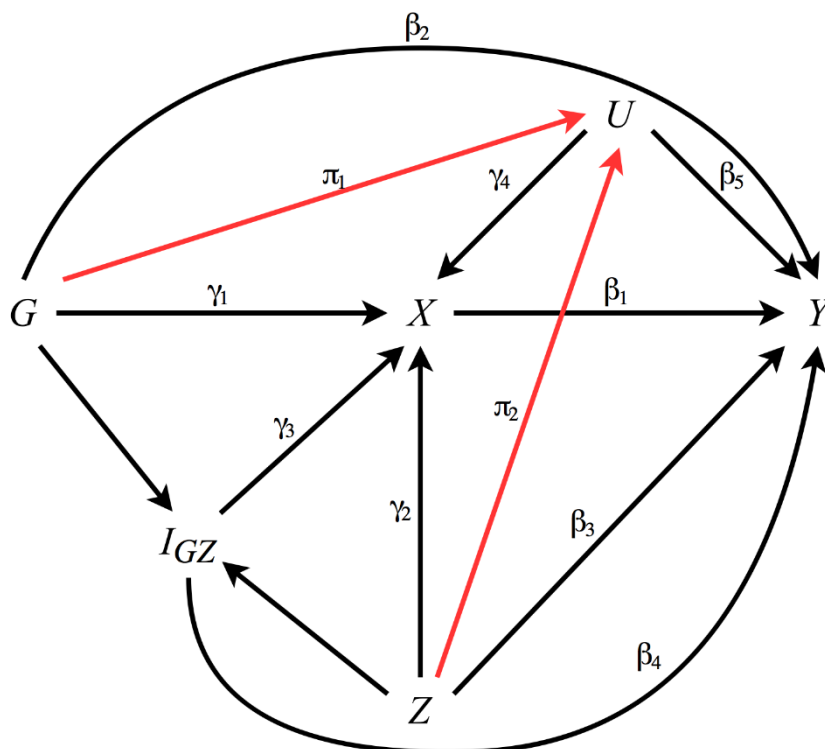
The corresponding Wald estimand is then given as:

$$\beta_1 + \frac{\beta_2\theta_1\pi_1 + \beta_3 + 2\beta_4\theta_1\pi_1Z_i + \beta_5\pi_1}{\gamma_1\theta_1\pi_1 + \gamma_2 + 2\gamma_3\theta_1\pi_1Z_i + \gamma_4\pi_1} \quad (\text{A22-10})$$

In this case term  $2\beta_4\theta_1\pi_1Z_i$  has an equivalent role in invalidating MRGXE estimates as the

$2\gamma_3\theta_1\pi_1Z_i$  term from scenario 1.

**Scenario 3:**



**Figure A7:** DAG illustrating a case in which the interaction-covariate  $Z$  and the instrument  $G$  are determinants of the confounder  $U$ .

Data generating model

$$G_i = N(0,1) \quad (\text{A23-1})$$

$$Z_i = N(0,1) \quad (\text{A23-2})$$

$$U_i = \pi_0 + \pi_1 G_i + \pi_2 Z_i + \epsilon_U \quad (\text{A23-3})$$

$$X_i = \gamma_0 + \gamma_1 G_i + \gamma_2 Z_i + \gamma_3 GZ_i + \gamma_4 U + \epsilon_X \quad (\text{A23-4})$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 G_i + \beta_3 Z_i + \beta_4 GZ_i + \beta_5 U_i + \epsilon_Y \quad (\text{A23-5})$$

Definition of bias term

First a model for the exposure  $X$  in terms of  $G$  is constructed by substituting models (A23-2) and (A23-3) into model (A23-4):

$$X_i = G_i(\gamma_1 + \gamma_3 Z_i + \gamma_4 \pi_1) + \eta_X \quad (\text{A23-6})$$

We can also construct a model for the outcome  $Y$  in terms of  $G$  by substituting models (A23-2), (A23-3), and (A23-6) into model (A23-5):

$$Y_i = G_i(\beta_1(\gamma_1 + \gamma_3 Z_i + \gamma_4 \pi_1) + \beta_2 + \beta_4 Z_i + \beta_5 \pi_1) + \eta_Y \quad (\text{A23-7})$$

Using models (A23-6) and (A23-7), we can calculate the partial effect of  $G$  for the first and second stage models respectively:

$$\frac{dX}{dG} = \gamma_1 + \gamma_3 Z_i + \gamma_4 \pi_1 \quad (\text{A23-8})$$

$$\frac{dY}{dG} = \beta_1(\gamma_1 + \gamma_3 Z_i + \gamma_4 \pi_1) + \beta_2 + \beta_4 Z_i + \beta_5 \pi_1 \quad (\text{A23-9})$$

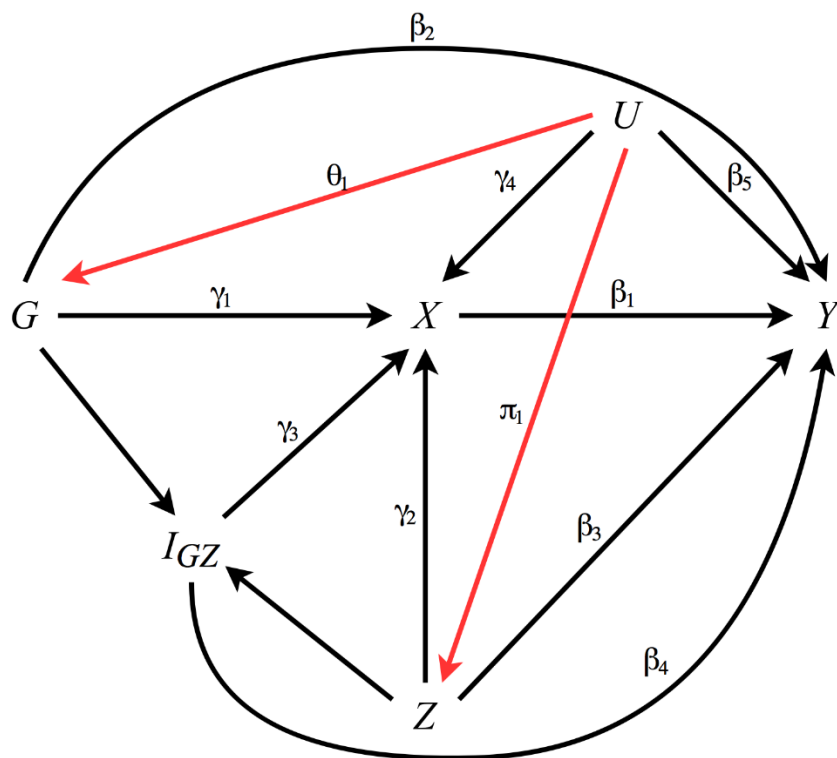
The corresponding Wald estimand is then given as:

$$\beta_1 + \frac{\beta_2 + \beta_4 Z_i + \beta_5 \pi_1}{\gamma_1 + \gamma_3 Z_i + \gamma_4 \pi_1} \quad (\text{A23-10})$$

As can be seen from equation (A23-10), the effect of  $G$  through the confounder  $U$  is constant and incorporated into the MRGXE pleiotropy correction. As a consequence, MRGXE provides an unbiased estimate of the causal effect  $\beta_1$ .



**Scenario 4:**



**Figure A8:** DAG illustrating a case in which the confounder  $U$  is a determinant of both the interaction-covariate  $Z$  and the instrument  $G$ .

Data generating model

$$U_i = N(0,1) \quad (\text{A24-1})$$

$$Z_i = \pi_0 + \pi_1 U_i + \epsilon_Z \quad (\text{A24-2})$$

$$G_i = \theta_0 + \theta_1 U_i + \epsilon_G \quad (\text{A24-3})$$

$$X_i = \gamma_0 + \gamma_1 G_i + \gamma_2 Z_i + \gamma_3 GZ_i + \gamma_4 U_i + \epsilon_X \quad (\text{A24-4})$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 G_i + \beta_3 Z_i + \beta_4 GZ_i + \beta_5 U_i + \epsilon_Y \quad (\text{A24-5})$$

Definition of bias term

We can substitute model (A24-2) into models (A24-4) and (A24-5), such that

$$X_i = U_i(\gamma_1\theta_1 + \gamma_2\pi_1 + \gamma_3\theta_1\pi_1 U_i + \gamma_4) + \eta_X \quad (\text{A24-6})$$

$$Y_i = U_i(\beta_1(\gamma_1\theta_1 + \gamma_2\pi_1 + 2\gamma_3\theta_1\pi_1 U_i + \gamma_4) + \beta_2\theta_1 + \beta_3\pi_1 + 2\beta_4\theta_1\pi_1 U_i + \beta_5) + \eta_Y$$

As the instrument  $G$  is determined by  $U$ , it is appropriate to calculate the partial effect of  $U$  for the first and second stage models respectively:

$$\frac{dX}{dU} = \gamma_1\theta_1 + \gamma_2\pi_1 + 2\gamma_3\theta_1\pi_1U_i + \gamma_4 \quad (\text{A24-8})$$

$$\frac{dY}{dU} = \beta_1(\gamma_1\theta_1 + \gamma_2\pi_1 + 2\gamma_3\theta_1\pi_1U_i + \gamma_4) + \beta_2\theta_1 + \beta_3\pi_1 + 2\beta_4\theta_1\pi_1U_i + \beta_5 \quad \text{A24- (9)}$$

The corresponding Wald estimand is then given as:

$$\beta_1 + \frac{\beta_2\theta_1 + \beta_3\pi_1 + 2\beta_4\theta_1\pi_1U_i + \beta_5}{\gamma_1\theta_1 + \gamma_2\pi_1 + 2\gamma_3\theta_1\pi_1U_i + \gamma_4} \quad (\text{A24-10})$$

As with scenario 1 and scenario 2, the term  $2\beta_4\theta_1\pi_1U_i$  encapsulates the bias resulting from change in pleiotropic effects across subgroups. As a result, the MRGXE model will be invalidated where there are one or more confounders causally downstream of both the instrument  $G$  and the interaction covariate  $Z$ , and where the average effect of the set of confounders is not balanced.

### A3: Code for performing analyses (using a preformatted data frame).

```
#Where pheno_data is an appropriately formatted data frame containing candidate
#interaction covariates, and xy_data is a formatted data frame containing the
#exposure and outcome of interest. The last column of pheno_data should contain
#values for the genetic instrument G, written in this example as pheno_data$SCORE.
```

```
pheno_data<-pheno_data[,-c(1,1691)]
```

```
phenovec<-length(names(pheno_data))-1
```

```
F_stat_pvalue<-rep(99,phenovec)
```

```
F_stat<-rep(12345,phenovec)
```

```
GxEB1<-rep(99,phenovec)
```

```
GxEse<-rep(99,phenovec)
```

```
GxEp<-rep(99,phenovec)
```

```
GxE_lb<-rep(99,phenovec)
```

```
GxE_ub<-rep(99,phenovec)
```

```
summary(ivreg(xy_data$f.4080.0.0~xy_data$f.845.0.0|pheno_data$SCORE),diagnostics=T
)
```

```
summary(ivreg(xy_data$f.4080.0.0~xy_data$f.845.0.0|pheno_data$SCORE),diagnostics=T
)$coefficients[2,1]
```

```
summary(ivreg(xy_data$f.4080.0.0~xy_data$f.845.0.0|pheno_data$SCORE),diagnostics=T
)$coefficients[2,2]
```

```
summary(ivreg(xy_data$f.4080.0.0~xy_data$f.845.0.0|pheno_data$SCORE),diagnostics=T
)$coefficients[2,4]
```

```
for(i in 1:phenovec){
```

```
  GZ<-pheno_data[,1690] * pheno_data[,i]
```

```
  formula_temp<-as.formula(paste(c(paste("xy_data[,2]",
  paste(c("xy_data[,1]",paste(c("pheno_data",names(pheno_data)[i]),collapse="$"),"ph
  eno_data[,1690]")),collapse = " + ")), sep = " ~ ")
```

```
  ,paste(c(paste(c("pheno_data",names(pheno_data)[i]),collapse="$"),"pheno_data[,169
  0]","GZ")),collapse = " + ")),collapse = "|")
```

```
  F_stat_pvalue[i]<-summary(ivreg(formula_temp),diagnostics=T)$diagnostics[1,4]
```

```
  F_stat[i]<-summary(ivreg(formula_temp),diagnostics=T)$diagnostics[1,3]
```

```
  GxEB1[i]<-summary(ivreg(formula_temp),diagnostics=T)$coefficients[2,1]
```

```
  GxEse[i]<-summary(ivreg(formula_temp),diagnostics=T)$coefficients[2,2]
```

```
GxEp[i]<-summary(ivreg(formula_temp),diagnostics=T)$coefficients[2,4]
GxE_lb[i]<-confint(ivreg(formula_temp))[2,1]
GxE_ub[i]<-confint(ivreg(formula_temp))[2,2]
}

output<-
data.frame(names(pheno_data)[1:1689],F_stat,F_stat_pvalue,GxEB1,GxEse,GxEp,GxE_lb,
GxE_ub)

write.csv(output,"output_extended.csv",row.names=F)
```

Further specific information pertaining to code is available upon request.

