1. **Model and likelihood**

The proposed model is

(1)

Where the equation (1) is for the gene expression data and the equation (2) is for the GWAS data. Here, and are the intercepts for the two models, respectively; is a *p*-vector of instrumental effect sizes on the explanatory variable; is a scalar that represents the causal effect of the explanatory variable on the outcome variable; is a *p*-vector of horizontal pleiotropic effect sizes of *p* instruments on the outcome variable; is an -vector of residual error with each element independently and identically distributed from a normal distribution ;and is an -vector of residual error with each element independently and identically distributed from a normal distribution . We note that while the above two equations are specified based on two separate studies, they are joined together with the common parameter . We assume .

Given , we have

Given , the observed data are and the observed likelihood function is

the last equality is due to that given , and are independent.

While ()

Where ,

Let

The above formula is equal to

where

Note that the last term is a Gaussian kernel

Thus

Let indicate all model parameters.

The hypothesis test for is

The likelihood ratio test (LRT) is given by

Where is the parameter estimator, and is the estimator under . Similarly, the hypothesis test for is

The LRT is given by

Where is the parameter estimator under .

1. **Estimation procedure**

We develop an expectation-maximization (EM) algorithm for inference, where we treat the SNP effect sizes as missing data. Traditional EM algorithm converges very slowly while Newton’s method may be unstable and sensitive to initial values. Therefore, we use a parameter-expanded version of EM, i.e. PX-EM[1](#_ENREF_1), for estimation. PX-EM improves the convergence rate of traditional EM algorithm while is simple to implement and enjoys the stability of traditional EM. To do so, we consider the parameter expanded version of our model as follows

(3)

(4)

Where is the expanded parameter. Let denote all parameters for parameter expanded model. When , the expanded model is equal to the model for the observated data. The reduction function can be defined as .

From the derivation similar as above, it is easy to obtain that, given **, ,**  and, the distribution of the latent variable is a normal distribution , where

The complete likelihood for the parameter expanded model can be calculated as

()

**In the E-step**, we derive function by taking expectation of the complete-data log-likelihood with respect to the distribution. Remember thatfor any symmetric matrix *A*, where denotes the trace of matrix *M*. We can get

Given the current value and the observed data, the function is

**In the M-step**, by setting the derivative of function to zero, we obtain

the new updates for all parameters. Where

Where denote the length *d* vector with all elements to be 1.

**In the Reduction step**, we re-set the estimation of parameters using the reduction function , and re-assigned .

Finally, for TWAS applications, we note that some genes have close to zero heritability. For genes whose expression levels are not affected by cis-SNP genotypes, not much information is available for the estimation of , which subsequently leads to an extremely large standard error of . Therefore, PMR-Egger can return the p value equal or close to 1 for these genes.

1. **Causal effect identification**

The causal interpretation of the parameter and its identification can be derived under the framework of decision-theoretic causal inference[2-5](#_ENREF_2). We define the causal effect of gene expression on the phenotype as the difference between the expected values of under an intervention that imposes on a reference value and another intervention that imposes another value . Let the symbol label the regime under which the value of is generated, with indicating that is fixed to value by an intervention, and denoting the observational regime under which the data have actually been generated. Then the average causal effect (ACE) of on the continuous phenotype is defined by

Let the notation indicates that is independent of given .

Our proposed MR model based on the observational data obtained under has been presented in Figure S1. Note that we directly model the horizontal pleiotropic effects through the arrow. Therefore, our model does not require the Exclusion Restriction condition of traditional MR. However, the other two assumptions in tradtional MR must be satisfied; that is, is associated with , and **(1)**.

One must note that requirements relating only to the observational regime can never be sufficient to estimate the causal effect of on , which is defined in terms of interventional regimes. Instead, we need to make additional assumptions that relate the observational regime to the interventional regimes . Under the assumption that the unobserved is a sufficient covariate for the effect of on , we can do this by elaborating Figure S1 to explicitly include the nonstochastic regime indicator for , as in the following Figure. For , this recovers the assumptions embedded in Figure S1, but in addition it relates the observational structure to what would happen under an intervention to set .

Figure. The causal diagram of PMR-Egger with the nonstochastic regime indicator

Z

*X*

U

y

F*X*

α

γ

It is illustrated that an intervention on will not affect or U, that is  **(2)**. And conditional on and U, the distribution of given does not depend on whether the value of has been generated by passive observation or intervention; that is **(3).** Furthermore, the formula **(1)** can be extended to **(4)**.

We can describe the dependence of on (the same in all regimes by (3)) by a linear

model: **(5)**, where is some function of .

Because (5) holds in the interventional regime , we deduce

where and is a constant independent of following (2). Thus can be interpreted causally, as it describs how the mean of responds to manipulation of *X*. Next we show how to estimate .

Again by (3), the formula (5) is also . Then

By (4), the first term on the right side is constant, thus

**(6)**

Equation (6) relates two functions of , each of which can be identified from observational

data. Consequently, we can estimate the causal parameter from such data.

1. **PMR-Egger model for summary statistics**

we denote the LDstructure of the cis-SNPs for one specific gene as in gene expression data, and in GWAS data; both are symmetric positive definite matrices. The marginal estimates of on ***x*** are (eQTL effects), the marginal estimates of on are (GWAS effects). While the corresponding conditional estimates are , .

The corresponding model for summary statistics are

Where ,,is the pleiotropy effect vector with each element equal to common parameter .

Given , and are independent.

The observed likelihood function is

Where ,

Let

The above formula is equal to

where =. Thus

Let indicate all model parameters.

The hypothesis test for is

The likelihood ratio test (LRT) is given by

Where is the parameter estimator, and is the estimator under . Similarly, the hypothesis test for is

The LRT is given by

Where is the parameter estimator under .

1. **PX-EM algorithm for summary statistics**

The parameter expanded version of our model is

Where is the expanded scalar parameter. Let denote all parameters for parameter expanded model. When , the expanded model is equal to the model for the observated data. The reduction function can be defined as .

Treating as the missing variable, the complete data likelihood is

Given the data and , and parameters, it is easy to obtain the distribution of is Gaussian with the mean

the covariance matrix

Using formula for any symmetrix *A.*

The *E*-step can generate the following function

In the *M*-step, we set the derivative of the function to zero and obtain the following updated equations for all parameters.

where is a length p vector with elements equal to 1.

**Reference**

1. Liu, C., Rubin, D.B., and Wu, Y.N. (1998). Parameter expansion to accelerate EM: the PX-EM algorithm. Biometrika 85, 755-770.

2. Berzuini, C., Guo, H., Burgess, S., and Bernardinelli, L. (2018). A Bayesian approach to Mendelian randomization with multiple pleiotropic variants.

3. Berzuini, C., Dawid, P., and Bernardinell, L. (2012). Causality: Statistical perspectives and applications.(John Wiley & Sons).

4. Dawid, A.P. (2000). Causal inference without counterfactuals. Journal of the American statistical Association 95, 407-424.

5. Dawid, A.P. (2015). Statistical causality from a decision-theoretic perspective. Annual Review of Statistics and Its Application 2, 273-303.