ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients

Seongho Kim[1,a]

[a]Biostatistics Core, Karmanos Cancer Institute, Wayne State University, USA

Abstract

Lack of a general matrix formula hampers implementation of the semi-partial correlation, also known as part correlation, to the higher-order coefficient. This is because the higher-order semi-partial correlation calculation using a recursive formula requires an enormous number of recursive calculations to obtain the correlation coefficients. To resolve this difficulty, we derive a general matrix formula of the semi-partial correlation for fast computation. The semi-partial correlations are then implemented on an R package **ppcor** along with the partial correlation. Owing to the general matrix formulas, users can readily calculate the coefficients of both partial and semi-partial correlations without computational burden. The package **ppcor** further provides users with the level of the statistical significance with its test statistic.

Keywords: correlation, partial correlation, part correlation, ppcor, semi-partial correlation

1. Introduction

The partial and semi-partial (also known as part) correlations are used to express the specific portion of variance explained by eliminating the effect of other variables when assessing the correlation between two variables (James, 2002; Johnson and Wichern, 2002; Whittaker, 1990). The great number of studies have been published using either partial or semi-partial correlations in many areas including cognitive psychology (e.g., Baum and Rude, 2013), genomics (e.g., Fang *et al.*, 2009; Kim and Yi, 2007; Zhu and Zhang, 2009), medicine (e.g., Vanderlinden *et al.*, 2013), and metabolomics (e.g., Kim *et al.*, 2012; Kim and Zhang, 2013).

The partial correlation can be explained as the association between two random variables after eliminating the effect of all other random variables, while the semi-partial correlation eliminates the effect of a fraction of other random variables, for instance, removing the effect of all other random variables from just one of two interesting random variables. The rationale for the partial and semi-partial correlations is to estimate a direct relationship or association between two random variables. The brief explanation follows to describe the main difference among the correlation, the partial correlation and the semi-partial correlation. Suppose there are three random variables (or vectors), X, Y and Z, and we are interested in the relationship (or association) between X and Y, for illustration purposes. Three situations are taken into consideration as shown in Figure 1. The figure describes three cases that (a) Z is correlated with none of X and Y, (b) only the random variable Y is correlated
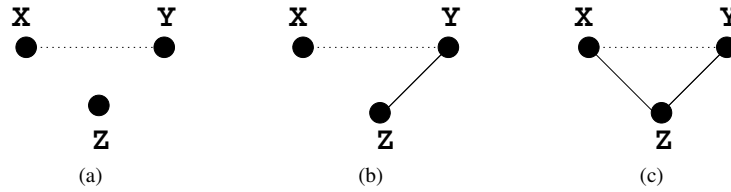
Figure 1: *Graphical illustration of partial and semi-partial correlations among the three random variables X, Y, and Z.*

with Z and (c) Z is correlated with both X and Y. Because Z is independent of both X and Y in Figure 1(a), the correlation, partial correlation and semi-partial correlation all should theoretically have the identical value. When only Y is correlated with Z as shown in Figure 1(b), the partial correlation is exactly same as the semi-partial correlation, but is different from the correlation. However, in case of Figure 1(c), all three correlations are different from each other. For more details on the partial and semi-partial correlations, refer to James (2002) and Whittaker (1990).

Several R packages have been developed only for the partial correlation. The R package **corpcor** (Schafer and Strimmer, 2005a) provides the function `cor2pcor()` for computing partial correlation from correlation matrix and vice versa. The function `ggm.estimate.pcor()`, which is built in the package **GeneNet** (Schafer and Strimmer, 2005b), allows users to estimate the partial correlation for graphical Gaussian models. The package **PCIT** (Watson-Haigh *et al.*, 2010) provides an algorithm for calculating the partial correlation coefficient with information theory. The function `partial.cor()` is included in **Rcmdr** (Fox, 2005) package for computing partial correlations. The package **parcor** (Kramer *et al.*, 2009) can be used for regularized estimation of partial correlation matrices. The **qp** (Castelo and Roverato, 2006) package provides users with q-order partial correlation graph search algorithm. The R package **space** (Peng *et al.*, 2009) can be used for sparse partial correlation estimation. However, none of these packages provide the level of significance for the partial correlation coefficient such as `p-value` and `statistic`. Furthermore, to our knowledge, there exists no R package for semi-partial correlation calculation.

On the other hand, there is no attempt to reduce the computational burden of the higher-order semi-partial correlation coefficients, while the higher-order partial correlation coefficients can be easily calculated using the inverse variance-covariance matrix. This means that a recursive formula (e.g., see Equation (2.2)) should be used for the higher-order semi-partial correlation calculation, hampering the use of the semi-partial correlation for high-dimensional data, such as 'omics' data, due to its expensive computation.

For these reasons, we derive a general matrix formula for the semi-partial correlation calculation (see Equation (2.6)). Using this general matrix formula, the semi-partial correlation coefficient can be simple but fast calculated. In order for the partial and the semi-partial correlations to be used practically, an R package **ppcor** is further developed in the R system for statistical computing (R Development Core Team, 2015). It provides a means for fast computing partial and semi-partial correlation as well as the level of statistical significance. The package **ppcor** is publicly available from CRAN at http://CRAN.r-project.org/package=ppcor and it is also available in the Supplementary Material at CSAM homepage (http://csam.or.kr).

2. Partial and Semi-partial Correlations

Consider the random vector $X = (x_1, x_2, \ldots, x_i, \ldots, x_n)'$ where $|X| = n$. We denote the variance of a random variable x_i and the covariance between two random variables x_i and x_j as $v_i \ (= \text{var}(x_i))$

and c_{ij} (= cov(x_i, x_j)), respectively. The variance-covariance matrices of random vectors X and X_S ($S \subset \{1, 2, \ldots, n\}$ and $|S| < n$) are denoted by C_X and C_S, respectively, where X_S is a random sub-vector of the random vector X. The correlation between two random variables x_i and x_j is denoted by $r_{ij} = c_{ij}/(\sqrt{v_i} \sqrt{v_j})$ (= corr(x_i, x_j)).

Definition 1. *The partial correlation of x_i and x_j given x_k is*

$$r_{ij|k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{1 - r_{ik}^2} \sqrt{1 - r_{jk}^2}} \tag{2.1}$$

and the semi-partial correlation of x_i with x_j given x_k is

$$r_{i(j|k)} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{1 - r_{jk}^2}}. \tag{2.2}$$

Whittaker (1990) defined the partial correlation using the correlation between two residuals. In fact, we can easily see that the definition in Whittaker (1990) is equivalent to the definition in Equation (2.1). Using the above definition, we can readily obtain another version of the definitions of the partial and the semi-partial correlation coefficients as follows.

Corollary 1. *The partial correlation of x_i and x_j given x_k, $r_{ij|k}$, is equal to*

$$corr(resid(i|k), resid(j|k)) = \frac{c_{ij} - c_{ik}v_k^{-1}c_{kj}}{\sqrt{v_i - c_{ik}v_k^{-1}c_{ki}} \sqrt{v_j - c_{jk}v_k^{-1}c_{kj}}} \tag{2.3}$$

and the semi-partial correlation of x_i with x_j given x_k, $r_{i(j|k)}$, is equal to

$$r_{ij|k} \frac{\sqrt{var(resid(i|k))}}{\sqrt{v_i}} = \frac{c_{ij} - c_{ik}v_k^{-1}c_{kj}}{\sqrt{v_i} \sqrt{v_j - c_{jk}v_k^{-1}c_{kj}}}, \tag{2.4}$$

where $resid(i|k) = x_i - \hat{x}_i(x_k)$ and $\hat{x}_i(x_k) = c_{ik}v_k^{-1}x_k$.

Note that the proof of Corollary 1 is omitted since it is straightforward. Corollary 1 can be further generalized to the case that there are two or more given variables. In other words, it can be extended to the higher-order partial and the higher-order semi-partial correlations. To do this, we need to consider the inverse variance-covariance matrix of X, $D_X = C_X^{-1}$. For the simplicity's sake, we denote D_X as $[d_{ij}]$ and C_X as $[c_{ij}]$, where d_{ij} and c_{ij} are the $(i, j)^{th}$ cell of the matrices D_X and C_X, respectively, and $1 \leq i, j \leq n$. Then the partial correlation between two variables given a set of variables can be calculated by the following lemma.

Lemma 1. (Whittaker, 1990) *The partial correlation of x_i and x_j given a random vector X_S (= $X_{(i,j)}$), $r_{ij|S}$, is equal to*

$$-\frac{d_{ij}}{\sqrt{d_{ii}} \sqrt{d_{jj}}}, \tag{2.5}$$

where X_S (= $X_{(i,j)}$) is the random sub-vector of X after removing the random variables x_i and x_j and its size is $|S|$ (= $|X| - 2$).

Most R packages for calculation of the partial correlation use the matrix-based calculation which is based on Equation (2.5), since this method is much less computationally expensive than the method based on Equation (2.1). In this case, we have to calculate the inverse variance-covariance matrix D_X in order to obtain the partial correlations for all pairs of the random variables of X. Fortunately, it can be easily obtained with a simple code in R. For example, the partial correlation of x_i and x_j given $X_{(i,j)}$ of X is the $(i, j)^{th}$ cell of the following matrix:

```
R> -cov2cor(solve(cov(X)))
```

However, there is no matrix-based mathematical formula for the semi-partial correlation. Without a general matrix formula, users have to calculate the higher-order semi-partial correlation through a recursive formula in Equation (2.2), which is time-consuming for high-dimensional data. Therefore, it is highly desirable to have a general matrix formula for the fast higher-order semi-partial correlation calculation. In the next theorem, we drive a matrix-based mathematical formula for the semi-partial correlation calculation.

Theorem 1. *The semi-partial correlation of x_i with x_j given a random vector X_S (= $X_{(i,j)}$), $r_{i(j|S)}$, is equal to*

$$r_{ij|S} \frac{1}{\sqrt{d_{ii} - d_{ij}d_{jj}^{-1}d_{ji}}} \frac{1}{c_{ii}}. \tag{2.6}$$

Proof: By Equation (2.4), we have the following equation

$$r_{i(j|S)} = r_{ij|S} \frac{\sqrt{\text{var}(\text{resid}(i|S))}}{\sqrt{v_i}}. \tag{2.7}$$

Then, using Equation (2.7) and Lemma 1, we have

$$
\begin{aligned}
r_{i(j|S)} &= -\frac{d_{ij}}{\sqrt{d_{ii}}} \frac{\sqrt{|C_X|}}{\sqrt{|C_S|}\sqrt{c_{ii}}} \\
&= -\frac{d_{ij}}{\sqrt{d_{ii}}\sqrt{d_{jj}}} \frac{1}{\sqrt{d_{ii} - d_{ij}d_{jj}^{-1}d_{ji}}} \frac{1}{c_{ii}} \\
&= r_{ij|S} \frac{1}{\sqrt{d_{ii} - d_{ij}d_{jj}^{-1}d_{ji}}} \frac{1}{c_{ii}}.
\end{aligned}
$$

$\square$

Using Theorem 1, we can readily calculate the semi-partial correlation using several lines of an R code. For example, the semi-partial correlation of x_i with x_j given $X_{(i,j)}$ is the $(i, j)^{th}$ cell of the matrix obtained in the last line of the following code.

```
R> cx <- cov(X)
R> dx <- solve(cx)
```

```
R> pc <- -cov2cor(dx)
R> diag(pc) <- 1
R> pc/sqrt(diag(cx))/sqrt(abs(diag(dx)-t(t(dx^2)/diag(dx))))
```

It first calculates the variance-covariance matrix of X, which is `cx`, and its inverse variance-covariance matrix, which is `dx`. Then the semi-partial correlations are obtained using the partial correlations, which is `pc`, in the R code above. Note that when the determinant of variance-covariance matrix is numerically zero, the R package **ppcor** computes its pseudo-inverse using the Moore-Penrose generalized matrix inverse (Penrose, 1995). However, in this case, no `statistics` and `p-values` are provided if the number of variables is greater than or equal to the sample size.

While, to our knowledge, no R packages provide the level of statistical significance for partial correlation coefficient, the R package **ppcor** includes the calculation of `statistics` and `p-values` of each correlation coefficient for both partial and semi-partial correlations. Moreover, **ppcor** provides users with nonparametric partial and semi-partial correlation coefficients based on Kendall's and Spearman's rank correlations.

The statistics $t_{ij|S}$ and $t_{i(j|S)}$ of the partial and semi-partial correlation of x_i and (with) x_j given x_S ($= x_{(i,j)}$) are calculated, based on the works in Weatherburn (1968) and Sheskin (2003), by

$$t_{ij|S} = r_{ij|S} \sqrt{\frac{N-2-g}{1-r_{ij|S}^2}}; \qquad t_{i(j|S)} = r_{i(j|S)} \sqrt{\frac{N-2-g}{1-r_{i(j|S)}^2}}, \tag{2.8}$$

where N is the sample size and g is the total number of given (or controlled) variables. Using Equation (2.8), their `p-values` are calculated by

$$p_{ij|S} = 2\Phi_t\left(-\left|t_{ij|S}\right|, N-2-g\right); \qquad p_{i(j|S)} = 2\Phi_t\left(-\left|t_{i(j|S)}\right|, N-2-g\right), \tag{2.9}$$

where $\Phi_t(\cdot)$ is the cumulative density function of a Student's t distribution with the degree of freedom $N-2-g$. It is known that the standard error is $1/\sqrt{N-2-g}$ (Olkin and Finn, 1995; Stanley and Doucouliagos, 2012; Sharma, 2012).

In case of Kendall's rank correlation, the statistics are computed by (Abdi, 2007)

$$z_{ij|S} = r_{ij|S} \sqrt{\frac{9(N-g-2)(N-g+1)}{2(2N-2g+1)}}; \qquad z_{i(j|S)} = r_{i(j|S)} \sqrt{\frac{9(N-g-2)(N-g+1)}{2(2N-2g+1)}}. \tag{2.10}$$

Using Equations (2.10), their `p-values` are calculated by

$$p_{ij|S} = 2\Phi\left(-\left|z_{ij|S}\right|\right); \qquad p_{i(j|S)} = 2\Phi\left(-\left|z_{i(j|S)}\right|\right), \tag{2.11}$$

where $\Phi(\cdot)$ is the cumulative density function of a standard normal distribution. The standard error is $\sqrt{2(2N-2g+1)}/\sqrt{9(N-g-2)(N-g+1)}$ (Abdi, 2007).

3. Examples

The R package **ppcor** provides users with four functions which are `pcor()`, `pcor.test()`, `spcor()`, and `spcor.test()`. The function `pcor()` (`spcor()`) calculates the partial (semi-partial) correlations of all pairs of two random variables of a matrix or a data frame and provides the matrices of `statistics` and `p-values` of each pairwise partial (semi-partial) correlation. In order to compute

the pairwise partial (semi-partial) correlation coefficient of a pair of two random variables given one or more random variables, `pcor.test()` (`spcor.test()`) can be also used instead. We can see how to use these functions through the following examples. First the test data, `y.data`, need to be created after loading the package with the following R codes.

```
R> library(ppcor)
R> y.data <- data.frame(
+ hl = c(7,15,19,15,21,22,57,15,20,18),
+ disp = c(0,0.964,0,0,0.921,0,0,1.006,0,1.011),
+ deg = c(9,2,3,4,1,3,1,3,6,1),
+ BC = c(1.78e-02,1.05e-06,1.37e-05,7.18e-03,0,0,0,4.48e-03,2.10e-06,0)
+ )
```

This test data, `y.data`, consists of 10 samples from four variables, `hl`, `disp`, `deg`, and `BC`. This data set is available from Drummond *et al.* (2006) and Kim and Yi (2007). The original data cover the relationship between sequence and functional evolutions in yeast proteins. Here we look at only part of the large data for the illustrative purpose. Note that `hl`, `disp`, `deg`, and `BC` stand for half life, dispensability, degree, and betweenness-centrality, respectively. Please refer to Drummond *et al.* (2006) and Kim and Yi (2007) for more details.

We can then calculate all pairwise partial correlations of each pair of two variables given other variables with

```
R> pcor(x=y.data,method="spearman")
```

Then we obtain the following output:

```
$estimate
            hl       disp        deg         BC
hl     1.0000000 -0.7647345 -0.1367596 -0.7860646
disp -0.7647345  1.0000000 -0.4845966 -0.4506273
deg  -0.1367596 -0.4845966  1.0000000  0.4010940
BC   -0.7860646 -0.4506273  0.4010940  1.0000000


$p.value
             hl       disp        deg         BC
hl   0.00000000 0.02708081 0.7467551 0.02071908
disp 0.02708081 0.00000000 0.2236095 0.26248897
deg  0.74675508 0.22360945 0.0000000 0.32471409
BC   0.02071908 0.26248897 0.3247141 0.00000000


$statistic
            hl       disp        deg         BC
hl     0.0000000 -2.907150 -0.3381686 -3.114899
disp -2.9071501  0.000000 -1.3569947 -1.236464
deg  -0.3381686 -1.356995  0.0000000  1.072529
BC   -3.1148991 -1.236464  1.0725286  0.000000


$n
```

```
[1] 10

$gp
[1] 2

$method
[1] "spearman"
```

The output has six values, `estimate`, which is the partial correlation coefficient, `p-value`, which is the level of statistical significance, `statistic`, which is the test statistic for `p-value`, `n`, which is the total number of samples, `gp`, which is the number of given or controlled variables, and `method`, which is the used correlation method among Pearson's, Kendall's, and Spearman's correlation methods. In case that the users are interested in the partial correlation between `hl` and `disp` given `deg` and `BC`, we can compute the partial correlation with

```
R> pcor.test(x=y.data$hl,y=y.data$disp,z=y.data[,c("deg","BC")]
+ ,method="spearman")
```

Then we obtain the following output:

```
    estimate     p.value statistic  n gp    Method
1 -0.7647345 0.02708081  -2.90715 10  2 spearman
```

Similarly, the semi-partial correlations can be calculated with

```
R> spcor(x=y.data,method="spearman")
```

Then we obtain the following output:

```
$estimate
              hl       disp         deg          BC
hl     1.00000000 -0.4254609 -0.04949092 -0.4558649
disp  -0.59319449  1.0000000 -0.27689034 -0.2522965
deg   -0.06380762 -0.2560457  1.00000000  0.2023709
BC    -0.42262366 -0.1677612  0.14551866  1.0000000

$p.value
              hl       disp        deg         BC
hl     0.0000000 0.2933025 0.9073559 0.2562889
disp   0.1211334 0.0000000 0.5067562 0.5466351
deg    0.8806850 0.5404845 0.0000000 0.6307871
BC     0.2968811 0.6912998 0.7309799 0.0000000

$statistic
              hl       disp         deg         BC
hl     0.0000000 -1.1515898 -0.1213762 -1.2545787
disp  -1.8048658  0.0000000 -0.7058372 -0.6386584
deg   -0.1566153 -0.6488095  0.0000000  0.5061789
BC    -1.1422336 -0.4168368  0.3602815  0.0000000
```

```
$n
[1] 10

$gp
[1] 2

$method
[1] "spearman"
```

The semi-partial correlation of `hl` with `disp` given `deg` and `BC` is calculated with

```
R> spcor.test(x=y.data$hl,y=y.data$disp,z=y.data[,c("deg","BC")]
+ ,method="spearman")
```

Then we obtain the following output:

```
    estimate    p.value statistic  n gp   Method
1 -0.4254609 0.2933025  -1.15159 10  2 spearman
```

It should be noted that, if a general matrix formula for the semi-partial correlation is not available, users have to calculate all pairs of each variable with the function spcor.test using two loops. To see how fast the general matrix formula can compute the semi-partial correlation, we compared the computational time by generating a data matrix with the size of 500×100 (i.e., the number of variables is 100 and the number of samples 500). When the function `spcor()` used, the total amount of computation time was 0.02 second, while it took 135.33 second when the function `spcor.test()` used with two loops. It demonstrates that the general matrix formula dramatically reduce the computational burden of the higher-order semi-partial correlation calculation. Note that this simulation was implemented on a desktop with Intel Core 2 Duo CPU 3.00 GHz.

4. Conclusions

A general matrix formula for the semi-partial correlation is derived. Lack of this general matrix formula has hampered implantation of the higher-order semi-partial correlation for high-dimensional 'omics' data analysis because it requires an enormous number of recursive calculations to obtain the correlation coefficient when using a recursive formula in Equation (2.2). However, using the derived matrix formula in Theorem 1, we can clearly see that the higher-order semi-partial correlation coefficient is calculated as simple but fast as the partial correlation does. The developed R package **ppcor** further provides users not only with a function to readily calculate the higher-order both partial and semi-partial correlations but also with `statistics` and `p-values` of the correlation coefficients.

5. Computational Details

The results in this paper were obtained using R 3.2.2 with the package **ppcor**. R and the **ppcor** package are available from CRAN at "http://CRAN.R-Project.org/ and in the Supplementary Material at CSAM homepage (http://csam.or.kr). Note that, in this latest version of the package **ppcor**, the `p-values` for Pearsons and Spearmans correlations are calculated based on the t-distribution and the Moore-Penrose generalized inverse matrix will be used when variance-covariance matrix is singular.

6. Acknowledgements

References

Abdi, H. (2007). Kendall rank correlation, In N.J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics*, Thousand Oaks (CA), Sage, 508–510.

Baum, E. S. and Rude, S. S. (2013). Acceptance-enhanced expressive writing prevents symptoms in participants with low initial depression, *Cognitive Therapy and Research*, **37**, 35–42.

Castelo, R. and Roverato, A. (2006). A robust procedure for Gaussian graphical model search from microarray data with *p* larger than *n*, *Journal of Machine Learning Research*, **7**, 2621–2650.

Drummond, D. A., Raval, A. and Wilke, C. O. (2006). A single determinant dominates the rate of yeast protein evolution, *Molecular Biology and Evolution*, **23**, 327–337.

Fang, X. Z., Luo, L., Reveille, J. D. and Xiong, M. (2009). Discussion: Why do we test multiple traits in genetic association studies?, *Journal of the Korean Statistical Society*, **38**, 17–23.

Fox, J. (2005). The R Commander: A basic-statistics graphical user interface to R, *Journal of Statistical Software*, **14**, 1–42.

James, S. (2002). *Applied Multivariate Statistics for the Social Sciences*, Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Johnson, R. A. and Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*, Prentice Hall.

Kim, S., Koo, I., Jeong, J., Wu, S., Shi, X. and Zhang, X. (2012). Compound identification using partial and semipartial correlations for gas chromatography-mass spectrometry data, *Analytical Chemistry*, **12**, 6477–6487.

Kim, S. and Yi, S. (2007). Understanding relationship between sequence and functional evolution in yeast proteins, *Genetica*, **131**, 151–156.

Kim, S. and Zhang, X. (2013). Comparative analysis of mass spectral similarity measures on peak alignment for comprehensive two-dimensional gas chromatography mass spectrometry, *Computational and Mathematical Methods in Medicine*, **2013**, 509761.

Kramer, N., Schafer, J. and Boulesteix, A. L. (2009). Regularized estimation of large scale gene association networks using Gaussian graphical models, *BMC Bioinformatics*, **10**, 384.

Olkin, I. and Finn, J. D. (1995). Correlations redux, *Psychological Bulletin*, **118**, 155–164.

Peng, J., Wang, P., Zhou, N. and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models, *Journal of the American Statistical Association*, **104**, 735–746.

Penrose, R. (1995). A generalized inverse for matrices, In *Proceedings of the Cambridge Philosophical Society*, **51**, 406–413.

R Development Core Team (2015). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org/

Schafer, J. and Strimmer, K. (2005a). A shrinkage approach to large-scale covariance matrix estimation and implications for functional Genomics, *Statistical Applications in Genetics and Molecular Biology*, **4**, 32.

Schafer, J. and Strimmer, K. (2005b). An empirical Bayes approach to inferring large-scale gene association networks, *Bioinformatics*, **21**, 754–764.

Sharma, J. K. (2012). *Business Statistics*, Pearson Education India.

Sheskin, D. J. (2003). *Handbook of Parametric and Nonparametric Statistical Procedures: Third*

Sharma, J. K. (2012). *Business Statistics*, Pearson Education India.

Sheskin, D. J. (2003). *Handbook of Parametric and Nonparametric Statistical Procedures: Third Edition*, CRC Press.

Stanley, T. D. and Doucouliagos, H. (2012) *Meta-Regression Analysis in Economics and Business*, Routledge.

Vanderlinden, L. A., Saba, L. M., Kechris, K., Miles, M. F., Hoffman, P. L. and Tabakoff, B. (2013). Whole brain and brain regional coexpression network interactions associated with predisposition to alcohol consumption, *PLoS ONE*, **8**, e68878.

Watson-Haigh, N. S., Kadarmideen, H. N. and Reverter, A. (2010). **PCIT**: An R Package for weighted gene co-expression networks based on partial correlation and information theory approaches, *Bioinformatics*, **26**, 411–413.

Weatherburn, C. E. (1968). *A First Course Mathematical Statistics*, Cambridge.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, John Wiley & Sons, New York.

Zhu, W. S. and Zhang, H. P. (2009). Why do we test multiple traits in genetic association studies?, *Journal of the Korean Statistical Society*, **38**, 1–10.