

ORIGINAL RESEARCH REPORT

Improving on Adjusted R-Squared

Julian Karch

The amount of variance explained is widely reported for quantifying the model fit of a multiple linear regression model. The default adjusted R-squared estimator has the disadvantage of not being unbiased. The theoretically optimal Olkin-Pratt estimator is unbiased. Despite this, it is not being used due to being difficult to compute. In this paper, I present an algorithm for the exact and fast computation of the Olkin-Pratt estimator, which facilitates its use. I compare the Olkin-Pratt, the adjusted R-squared, and 18 alternative estimators using a simulation study. The metrics I use for comparison closely resemble established theoretical optimality properties. Importantly, the exact Olkin-Pratt estimator is shown to be optimal under the standard metric, which considers an estimator optimal if it has the least mean squared error among all unbiased estimators. Under the important alternative metric, which aims for the estimator with the lowest mean squared error, no optimal estimator could be identified. Based on these results, I provide careful recommendations on when to use which estimator, which first and foremost depends on the choice of which metric is deemed most appropriate. If such a choice is infeasible, I recommend using the exact Olkin-Pratt instead of the default adjusted R-squared estimator. To facilitate this, I provide the R package `altR2`, which implements the Olkin-Pratt estimator as well as all other estimators.

Keywords: multiple linear regression; R-squared; amount of variance explained; effect size

Introduction

Multiple linear regression is one of the most used analysis methods within psychological research. The R-Squared (R^2) value is commonly reported when performing multiple linear regression. It quantifies the proportion of variance of the dependent variable that can be accounted for by the regression model in the sample, which is commonly abbreviated as the proportion of variance explained. It is well known that the R^2 value systematically overestimates the amount of variance explained in the *population*, which is arguably the more relevant quantity. To estimate the amount of variance explained in the population, the so-called adjusted R^2 is typically used.

While standard linear regression software, such as SPSS and the `lm` function in R (R Core Team, 2018), and consequently the vast majority of psychological publications report only R^2 and adjusted R^2 , many alternative estimators for the proportion of variance explained in the population have been proposed (see, Raju, Bilgic, Edwards, & Fleer, 1997; Shieh, 2008; Yin & Fan, 2001, for overviews). This leads to the question which of these estimators should be used. There have already been multiple studies addressing this issue (most notably, Raju et al., 1997; Shieh, 2008; Yin & Fan, 2001) via a systematic comparison of estimators.

In this study, I extend these previous comparisons in three aspects. First, I add a new estimator. This estimator is an implementation of the Olkin-Pratt estimator (Olkin & Pratt, 1958). The Olkin-Pratt estimator has been shown to be optimal under the most prevalent optimality criterion used in statistics, which considers an estimator optimal if it always has least mean squared error (MSE) among all unbiased estimators (Olkin & Pratt, 1958). Despite this favorable property, it is not used and has not been included in any of the previous comparisons due to the difficulties associated with computing it. Here, I show that the Olkin-Pratt estimator can be computed relatively straightforwardly, which facilitates its use.

Second, similarly to Raju et al. (1997), and Shieh (2008), I evaluate the estimators using both MSE and bias as both are essential. However, in contrast to Raju et al. (1997), Shieh (2008), I do not combine MSE and bias in an informal way but, instead, as suggested by concepts established within theoretical statistics (Lehmann & Casella, 2003). I will provide the detailed explanation of this strategy in the Methods section. Importantly, this strategy avoids non-transparent, conflicting recommendations such as Raju et al. (1997) recommending the use of the standard adjusted R^2 estimator and Shieh (2008) the positive-part Pratt estimator despite both basing their conclusion on bias and MSE. It also leads to considering the optimality of estimators based on different perspectives, which weight the relative importance of bias and MSE differently. To obtain bias and MSE of the estimators, I perform a Monte-Carlo simulation study.

Third, to enable usage of the alternative estimators, I provide an R package (“altR2” on CRAN), which contains all estimators compared here. This also includes all estimators compared previously.

This paper is structured as follows. First, I introduce the compared estimators, the design of the simulation study, and the different perspectives used for evaluation. Then, I discuss which estimator is optimal under which perspective. I end by relating the results to previous studies and a recommendation of which estimators to use.

Method

Proportion of Variance Explained in the Population

Consider $i = 1, \dots, N$ observations of the dependent variable Y_i and the corresponding p predictors $[X_{i1}, \dots, X_{ip}]$. The statistical model associated with multiple linear regression is then

$$Y_i = \beta_0 + \sum_{k=1}^p X_{ik} \beta_k + \epsilon_i \tag{1}$$

where β_0, \dots, β_p are the regression coefficients and ϵ_i a normal random variable with constant variance σ_ϵ^2 representing the residual error.

It must be differentiated whether the predictors are assumed to be fixed or random. I assume that the predictors are random since I agree with Yin and Fan (2001), that, in psychology, random predictors are the norm. In particular, in line with Shieh (2008), Yin and Fan (2001), I assume that the predictors have a joint multivariate normal distribution. Under this assumption, the dependent variable Y_i is normally distributed with constant variance, which I will denote as total variance σ_Y^2 .

The true proportion of variance explained in the population is

$$\rho^2 = 1 - \frac{\sigma_\epsilon^2}{\sigma_Y^2} \tag{2}$$

This is known as squared multiple correlation ρ^2 in the statistical literature, which I will use from now on.

Estimators

Normal R^2

The normal R^2 replaces the error variance and the total variance by their maximum likelihood estimates $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_Y^2$. This results in $R^2 = 1 - \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_Y^2}$. For a given sample, the maximum likelihood estimates can be calculated using the total sums of squares SS_T as well as the residual sums of squares SS_R : $\hat{\sigma}_Y^2 = SS_T / N$ and $\hat{\sigma}_\epsilon^2 = SS_R / N$. By inserting into the definition of R^2 one obtains

$$R^2 = 1 - \frac{SS_R}{SS_T}$$

Adjusted R^2

The core idea of adjusted R^2 , is to replace the biased estimators with their unbiased counterparts. The unbiased estimates are $\tilde{\sigma}_\epsilon^2 = SS_R / (N - p - 1)$ and $\tilde{\sigma}_Y^2 = SS_T / (N - 1)$

This leads to the adjusted R^2 , which is called Ezekiel estimator in the statistical literature:

$$\text{Ezekiel: } \hat{\rho}_E^2(R^2) = 1 - \frac{N-1}{N-p-1} (1-R^2).$$

Somewhat counterintuitively, the Ezekiel estimator is not an unbiased estimator of the squared multiple correlation despite using unbiased estimators for the error and the total variance.

Variations of the Ezekiel Estimator

In the literature, slight variations of the Ezekiel estimator have been proposed, which I also include in the comparison. Due to space constraints, I only list them here. For a detailed description of the estimators including their derivation, the reader is referred to Raju et al. (1997), Yin and Fan (2001) as well as the references therein.

$$\text{Smith: } \hat{\rho}_S^2(R^2) = 1 - \frac{N}{N-p} (1-R^2)$$

$$\text{Wherry: } \hat{\rho}_W^2(R^2) = 1 - \frac{N-1}{N-p} (1-R^2)$$

Maximum Likelihood Estimator

The maximum likelihood estimator $\hat{\rho}_{ML}^2(R^2)$ introduced by Alf and Graf (2002) estimates the squared multiple correlation by finding the squared multiple correlation value that maximizes the likelihood of the observed data. I refer the reader to Alf and Graf (2002) for the computational details.

Olkin-Pratt Estimator

Olkin and Pratt (1958) proposed an estimator for the squared multiple correlation ρ^2 and showed that it is the unique uniformly minimum variance unbiased estimator. That is, no matter the circumstances, the Olkin-Pratt estimator always has the smallest MSE among all unbiased estimators. Thus, among the unbiased estimators, it can be considered optimal. No such strong theoretical justification exists for any other estimator.

Despite this favorable property, the Olkin-Pratt estimator has not been implemented in any software package, is not being used, and has not been included in any previous comparison. The reason for this is that it has been believed to be difficult to compute (Shieh, 2008) since its formula contains the hypergeometric function. The hypergeometric function ${}_2F_1(a, b; c; z)$ is difficult to compute as it is defined via an infinite series (see Appendix A for the full definition). The Olkin and Pratt estimator is

$$\hat{\rho}_{Op}^2(R^2) = 1 - \frac{N-3}{N-p-1} (1-R^2) {}_2F_1\left(1, 1; \frac{N-p+1}{2}; 1-R^2\right).$$

Due to the computation complexity of the hypergeometric function, different approximation approaches exist, which I will review in the following.

Approximations of the Olkin-Pratt Estimator

One approach to approximate the Olkin-Pratt estimator is to only use the first $K + 1$ addends of the infinite series of the hypergeometric function. This leads to a family of estimators of the form

$$\hat{\rho}_{OPK}^2(R^2) = 1 - \frac{N-3}{N-p-1} (1-R^2) \sum_{k=0}^K t_k$$

where t_k is the k th addend of the hypergeometric series (see Appendix A).

In the previous comparisons, the $\hat{\rho}_{OP1}^2$, $\hat{\rho}_{OP2}^2$ and $\hat{\rho}_{OP5}^2$ estimators have been included, which I also included in the comparison.

Other approximation approaches modify the $\hat{\rho}_{OP1}^2$ estimator to correct for the left out addends. These are:

$$\text{Pratt: } \hat{\rho}_p^2(R^2) = \frac{N-3}{N-p-1} (1-R^2) \left(1 + \frac{2(1-R^2)}{N-p-2.3} \right)$$

$$\text{Claudy: } \hat{\rho}_c^2(R^2) = \frac{N-4}{N-p-1} (1-R^2) \left(1 + \frac{2(1-R^2)}{N-p+1} \right)$$

For the rationales of these corrections, see Raju et al. (1997), Yin and Fan (2001) as well as the references therein.

New Estimator: Exact Olkin-Pratt Estimator

The existence of the approximations of the Olkin-Pratt estimator is based on the assumption that the Olkin-Pratt estimator is too difficult to compute, making it unusable in practice. In Appendix A, I show that this assumption is not true. In particular, I demonstrate that there are closed-form solutions for the hypergeometric function for all possible inputs required for the evaluation of the Olkin-Pratt estimator. These closed-form solutions allow the exact Olkin-Pratt estimator to be computed within milliseconds and thus facilitating its use in practice.

Positive-Part Estimators

All estimators introduced here, besides the normal R^2 and the maximum likelihood estimator, share one disadvantage. The estimated squared multiple correlation can be negative, while the population squared multiple correlation cannot. In practice, this problem is commonly addressed by assuming that the population squared multiple correlation is 0 if an estimator returns a negative value. Shieh (2008) formalized this practice by suggesting that so-called positive-part estimators should be considered, which return 0 if the respective estimator returns a negative value and otherwise return the original estimate. Shieh (2008) showed that this increases the bias, as expected, but decreases MSE. Consequently, I also included the positive-part versions of all estimators introduced here in the comparison.

Simulation Study

Overview Design

In the simulation study, I varied the factors, sample size N , number of predictors p , and population squared multiple correlation ρ^2 . Fisher (1928) showed that these are the only factors modifying the sampling distribution of R^2 and thus the bias and MSE of each estimator.

Sample Size

I investigated the sample sizes 10, 20, 30, 40, 50, 60, 100, 150. According to the estimates provided by Marszalek, Barber, Kohlhart, and Holmes (2011), the resulting range contains more than 75% of the sample sizes reported in psychological journals from 1995–2006. I chose to cover the sample sizes below 60 more densely because these sample sizes occurred more often and include more than 50% of the sample sizes reported in psychological journals.

Number of Predictors

As number of predictors p , I chose 2, 5 and 10. This is equivalent to the values used by one previous comparison (Shieh, 2008) and close to the values used by another one (Yin & Fan, 2001).

Squared Multiple Correlation

For the true squared multiple correlation ρ^2 , I included the values investigated in the previous comparison performed by Shieh (2008), 0.0, 0.1, 0.2, ..., 0.9 as they sample the space of possible values in a relatively dense matter. I additionally included two small values: 0.01 and 0.05 as small squared multiple correlation values are relatively common in psychology (see, for example, Karch et al., 2019).

Data Generation

In Appendix B as well as in the code (<https://doi.org/10.24433/CO.8023088.v3>), I describe how I simulated data for a given combination of sample size N , number of predictors p , and squared multiple correlation ρ^2 in detail. For generating the data, some parameters, like the residual variance, had to be set to fixed values. Importantly, the results are not influenced by those choices, as the sampling distribution of R^2 and thus the bias and MSE of each estimator are independent of those parameter values (Fisher, 1928).

Replications

Fully crossing all factors would lead to 288 design cells. However, those design cells for which the sample size N was not at least two bigger ($N \geq p + 2$) than the number of predictors p were removed since some estimators, for example, the Ezekiel estimator, require this. This resulted in 276 remaining design cells. For each design cell, I repeated the process of generating data and obtaining the estimates 100,000 times.

Evaluating the estimators

Uniformly Minimum Mean Squared Error Unbiased

The classical optimality criterion for point estimators is uniformly minimum variance unbiasedness (Lehmann & Casella, 2003, Chapter 2). For example, the standard

estimator of the regression coefficients (ordinary least squares) is uniformly minimum variance unbiased. A uniformly minimum variance unbiased estimator has the minimum MSE among all unbiased estimators in all circumstances. I will call this property uniformly minimum MSE unbiased to connect it more obviously to MSE. This property has been proven for the Olkin-Pratt estimator. Thus the Olkin-Pratt estimator is unbiased. Since all other estimators have a different expected value, it follows that they are biased.

Mostly, to validate the proposed implementation of the Olkin-Pratt estimator, I quantified whether an estimator was empirically unbiased, for a given sample size N , number of predictors p , squared multiple correlation ρ^2 condition, by using the one-sample t-test, with the null hypothesis of zero bias. I conclude that an estimator is overall empirically unbiased for a given sample size N , and number of predictors p condition if for no squared multiple correlation ρ^2 value, the null hypothesis was rejected. As significance threshold I chose $\alpha = .005$ as recommended by Benjamin et al., 2018. Note that an estimator might still be biased despite being empirically unbiased. Indeed, all estimators besides the Olkin-Pratt estimator are biased. However, if they are empirically unbiased, the bias is likely negligible.

Uniformly Lowest Mean Squared Error

It has been argued that considering only unbiased estimators is detrimental as biased estimators typically have lower MSE than unbiased estimators (Lehmann & Casella, 2003, Chapters 4 and 5). At the same time, MSE is well suited as the only metric for choosing between estimators as it trades off bias and variance. This argument leads to MSE being considered as the only metric (Lehmann & Casella, 2003, Chapters 4 and 5).

When using the MSE as the only metric, an optimal estimator has the lowest MSE across all conditions, in the case of this study, across all sample size N , number of predictors p , and squared multiple correlation ρ^2 combinations. In practice, such a uniformly best estimator does rarely exist (Lehmann & Casella, 2003, Chapters 4 and 5). To confirm that this is the case, I identified the estimator with the lowest MSE for each combination of design factors.

Mean Squared Error Dominance

An estimator A dominates an estimator B if it always has lower MSE. This is an essential relationship because if an estimator is dominated by another estimator, it should never be used (if considering only MSE). Consequently, I also report the dominance relationships to determine the set of admissible estimators (those which are not dominated).

Maximum Mean Squared Error

The first solution to the problem that the uniformly lowest MSE rarely exists is typically used in frequentist inference. The strategy is to consider the maximum MSE over all unknown parameters (in this study, only the squared multiple correlation ρ^2) as the metric for

choosing between estimators. Consequently, I also report the maximum MSE for all estimators.

Average Mean Squared Error

The Bayesian solution to the fact that no estimator exists with uniformly the lowest MSE is to consider a weighted average of the MSEs over all unknown parameter values. The crucial question is which prior to use for the weighting. In line with the common practice of using flat priors in psychology (Albers, Kiers, & van Ravenzwaaij, 2018), I average using a uniform prior over the squared multiple correlation ρ^2 .

A Note On Consistency

Note that all of the estimators are consistent. Thus for a fixed number of predictors p , they will all converge to the true squared multiple correlation value ρ^2 as the sample size N increases. Consequently, it is to be expected that the differences between the estimators are most pronounced for smaller sample sizes.

Results

In the supplementary material (CSV Files S1, and S2, and Tables S3-S25), I present bias and MSE for each design cell. Here, I will focus on summarizing those values using the different evaluation strategies just discussed.

Uniformly Minimum Mean Squared Error Unbiased

Table 1 shows which estimators were empirically unbiased and under what condition. The exact Olkin-Pratt estimator was the only estimator that was empirically unbiased across *all* conditions. Additionally, only approximations of the exact Olkin-Pratt estimator were empirically unbiased in *any* condition, namely: the Olkin-Pratt $K = 2$, $K = 5$, and the Pratt estimator.

To investigate the size of the bias, I exemplarily display the bias of the most popular Ezekiel estimator, the previously recommended Pratt estimator, and the new exact Olkin-Pratt estimator in a small ($N = 10$) and big sample ($N = 150$) condition in **Figures 1a** and **1b** respectively. In the small sample condition, both the Ezekiel and the Pratt estimator were substantially biased. The Pratt estimator up to .09 and the Ezekiel estimator up to .05. Both estimators consistently underestimated the squared multiple correlation ρ^2 . In the large sample condition, all estimators were virtually unbiased, as expected.

In summary, the results confirm that the newly proposed implementation of the Olkin-Pratt estimator is unbiased, whereas none of the other estimators is.

Uniformly Lowest Mean Squared Error

In Appendix C, the estimator with the lowest MSE for each combination of design factors is shown. As expected, no estimator had uniformly lowest MSE across all conditions. Instead, 6 out of the total 20 estimators were best in at least one condition, namely: the positive-part versions of the Ezekiel, Smith, Wherry, and Claudy estimators as well as the maximum likelihood and the normal R^2 estimators.

Table 1: Emperically Unbiased Estimators.

<i>N</i>	<i>p</i>		
	2	5	10
10.00	OPE	OPE	
20.00	OP5, OPE	OP5, OPE	OP5, OPE
30.00	OP5, OPE	OP5, OPE	OP2, OP5, OPE
40.00	OP5, OPE	OP2, OP5, OPE, P	OP5, OPE, P
50.00	OP2, OP5, OPE, P	OP2, OP5, OPE, P	OP5, OPE, P
60.00	OP2, OP5, OPE	OP2, OP5, OPE, P	OP2, OP5, OPE, P
100.00	OP2, OP5, OPE, P	OP2, OP5, OPE, P	OP2, OP5, OPE, P
150.00	OP2, OP5, OPE, P	OP2, OP5, OPE, P	OP2, OP5, OPE, P

Note: For a sample size *N*, number of predictors *p* combination, the estimators are displayed that were deemed unbiased for all squared multiple correlation values ρ^2 . OPE = Exact Olkin-Pratt; OP2 = Olkin-Pratt, *K* = 2; OP5 = Olkin-Pratt, *K* = 5; P = Pratt.

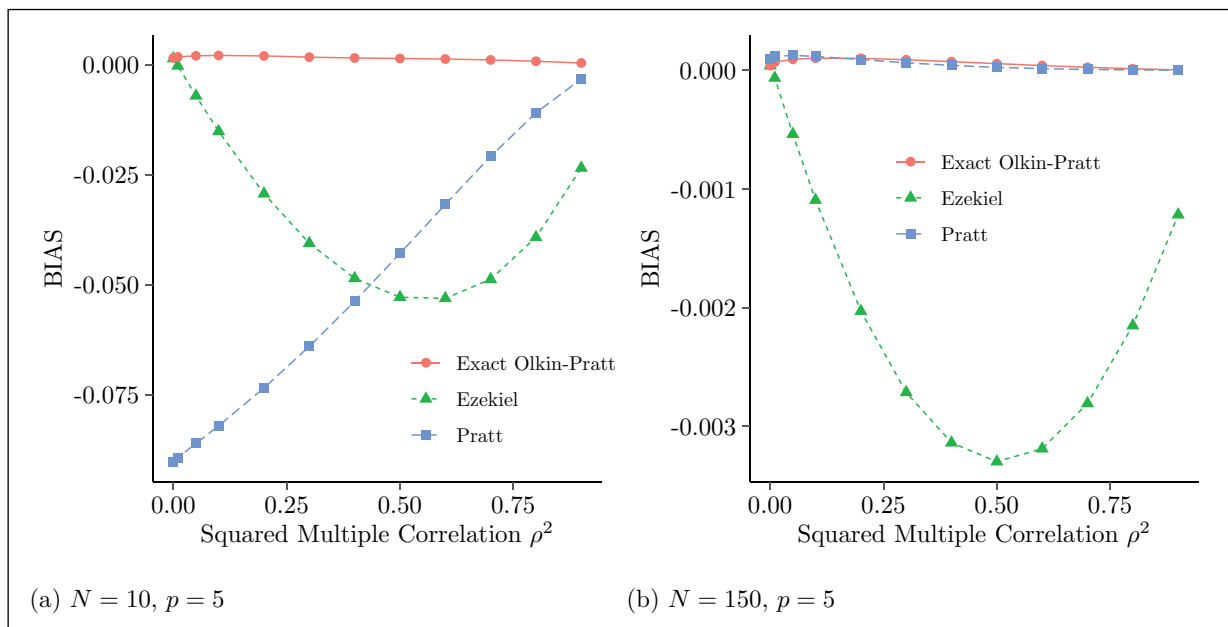


Figure 1: Bias for selected estimators in a small sample (a) and a big sample (b) condition.

In general, the unknown factor squared multiple correlation ρ^2 must be known in order to select the best estimator for all sample size *N*, number of predictor *p* combinations. The difference in terms of MSE between the estimators was non-trivial for some conditions, especially when the sample size was small, as is exemplarily displayed in **Figure 2a**. In contrast, for large sample sizes, the difference between the estimator was negligible, as expected (see **Figure 2b**). In summary, without knowledge of the true squared multiple correlation ρ^2 , it is impossible to choose the estimator with the lowest MSE, and there can be substantial differences in MSE between the estimators.

Dominance

While no estimator dominated all other estimators, multiple estimators dominated at least one other estimator. In **Figure 3**, the full dominance relationships between all estimators are presented.

The two most relevant results are as follows. First, all positive-part versions dominated their regular counterparts. Second, within the Olkin-Pratt estimators, the simpler versions dominated the more complex. This is reflected by the Olkin-Pratt *K* = *x* estimators dominating all Olkin-Pratt *K* = *y* estimators for which *x* < *y*, all Olkin-Pratt *K* estimators dominating the exact Olkin-Pratt estimator, and the Olkin-Pratt *K* = 1 estimator dominating the Pratt estimator.

The dominated estimators were not considered for the maximum MSE as well as the average MSE perspective since the estimators dominating them are guaranteed to also perform better with regard to these metrics.

Maximum Mean Squared Error

In **Table 2**, the estimator with the lowest maximum MSE for each sample size *N*, number of predictors *p* combination is displayed. No estimator was best in this

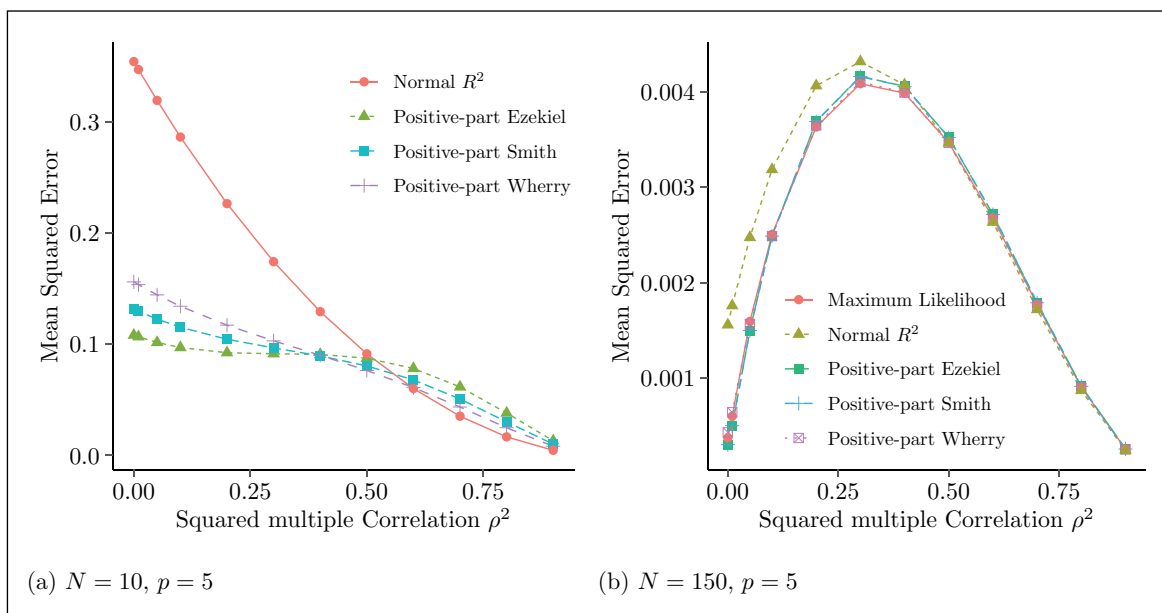


Figure 2: Mean squared error for selected estimators in a small sample (a) and a big sample (b) condition. For each condition, all estimators that had lowest mean squared error for a certain ρ^2 value were selected.

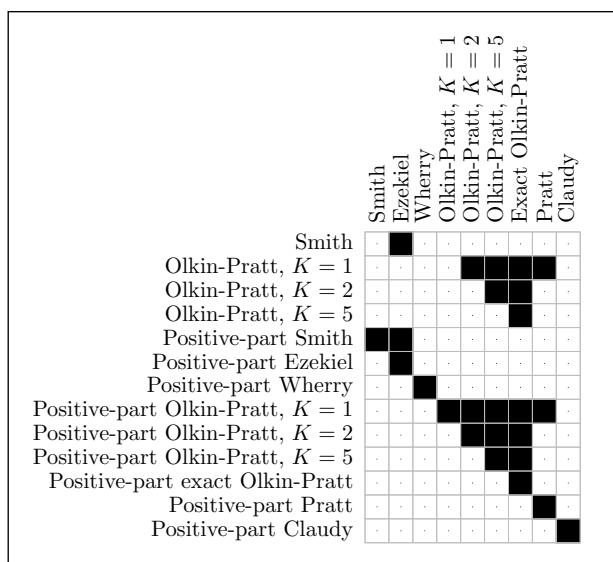


Figure 3: Matrix representing the dominance relationships between the estimators. A black square signals that the estimator in the respective row dominates the estimator in the corresponding column. Rows and columns that contain no black square have been removed.

Table 2: Estimators With the Lowest Maximum Mean Squared Error.

N	p							
	10	20	30	40	50	60	100	150
2	W ⁺	R ²	R ²	R ²	ML	ML	ML	ML
5	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	W ⁺	ML	ML
10		E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	ML	ML

Note: E⁺ = positive-part Ezekiel; ML = Maximum Likelihood; R² = normal R²; W⁺ = positive-part Wherry.

sense across all conditions. Instead, 4 out of 20 estimators were best, namely: the positive-part versions of the Ezekiel and the Wherry estimators as well as the maximum likelihood and the normal R² estimators.

To nevertheless obtain an overall best estimator, I also calculated the maximum MSE across all conditions for each estimator. The results are displayed in **Table 3**. Most importantly, the positive-part Ezekiel estimator was best with a maximum MSE of 0.1082. The maximum MSE of the second-best estimator (positive-part Pratt) was 0.1230 and thus more than 10% higher.

Average Mean Squared Error

In **Table 4**, the estimator with the lowest average MSE for each sample size N, number of predictors p combination is displayed. No estimator was best in this sense across all conditions. Instead, 4 out of 20 estimators were best, namely: the positive-part versions of the Ezekiel, Wherry, and the Smith estimators, as well as the maximum likelihood estimator.

To nevertheless obtain an overall best estimator, I also calculated the average MSE across all conditions for each estimator. The result is displayed in **Table 3**. Importantly, the positive-part Ezekiel estimator again was best with an average MSE of 0.015353. However, the average MSE of the second-best estimator (maximum likelihood) was 0.015369 and thus virtually identical.

Impact of Always Using the Positive-Part Ezekiel Estimator

Since the positive-part Ezekiel estimator was best both in terms of average as well as maximum MSE, I investigated the impact of always using it. To do this, I calculated the increase in maximum MSE and average MSE for each N, p combination instead of the optimal estimator.

Table 3: Maximum and Average Mean Squared Error.

	E⁺	P⁺	S⁺	ML	OPE⁺	OP5⁺	OP2⁺	OP1⁺	C⁺	W⁺	R²
Max. MSE	0.1082	0.1230	0.1317	0.1367	0.1380	0.1380	0.1387	0.1413	0.1533	0.1561	0.3543
Av. MSE	0.0154	0.0161	0.0154	0.0154	0.0161	0.0161	0.0161	0.0160	0.0166	0.0156	0.0260

Note: The minimum within each row is marked in bold. The average mean squared error was lowest for the E⁺ = estimator, when considering more than 4 digits.

E⁺ = Positive-part Ezekiel; P⁺ = positive-part Pratt; S⁺ = positive-part Smith; ML = Maximum Likelihood; OPE⁺ = positive-part exact Olkin-Pratt; OP2⁺ = positive-part Olkin-Pratt K = 2; OP5⁺ = positive-part Olkin-Pratt K = 5; C⁺ = positive-part Claudy; W⁺ = positive-part Wherry; R² = normal R².

Table 4: Estimators With The lowest Average Mean Squared Error.

<i>N</i>	<i>p</i>							
	10	20	30	40	50	60	100	150
2	W ⁺	W ⁺	W ⁺	W ⁺	ML	ML	ML	ML
5	E ⁺	S ⁺	W ⁺	W ⁺	W ⁺	W ⁺	ML	ML
10		E ⁺	S ⁺	S ⁺	ML	ML	ML	ML

Note: E⁺ = positive-part Ezekiel; S⁺ = positive-part Smith; W⁺ = positive-part Wherry; and ML = Maximum Likelihood.

The full results are displayed in **Table 5**. The increase in maximum MSE was never higher than 11%. The increase in mean MSE was never higher than 4%. Thus, the increase in MSE when always using the positive-part Ezekiel estimator was relatively mild.

Discussion and Conclusion

In this paper, I compared the novel exact Olkin-Pratt estimator and 19 additional estimators of the squared multiple correlation using different perspectives. These different perspectives all follow directly from optimality concepts established in theoretical statistics and are based on bias, MSE, or a combination thereof.

Regarding the most prevalent uniformly minimum MSE unbiased perspective, the results are unambiguous and in line with the theoretical optimality property established for the Olkin-Pratt estimator. The exact Olkin-Pratt estimator was optimal. It was the only estimator that was empirically unbiased across all conditions. Consequently, based on this perspective, the exact Olkin-Pratt estimator should always be used.

Regarding the perspectives that consider only MSE, the results are more ambiguous. No estimator had uniformly lowest MSE across all conditions. Even more importantly, no estimator was uniformly best according to the maximum or average MSE perspectives. However, across all conditions, the positive-part version of the most widely used Ezekiel (adjusted R²) estimator performed best both according to the maximum MSE as well as average MSE perspective.

To choose the best estimator under the MSE-only perspective, two cases have to be distinguished. First, there is some knowledge about the squared multiple correlation ρ². Second, this is not the case. If there is some knowledge, then I advise using Table C1 in Appendix C to select the estimator with the lowest MSE for the particular study. If

Table 5: Relative Increase of the Average and Maximum Mean Squared Error When Always Using the Positive-part Ezekiel Estimator.

<i>N</i>	Maximum; <i>p</i>			Average; <i>p</i>		
	2	5	10	2	5	10
10.00	8.63	0.00		2.14	0.00	
20.00	10.56	4.98	0.00	3.63	0.64	0.00
30.00	7.68	5.02	2.37	3.27	1.20	0.13
40.00	6.53	5.10	3.94	2.99	1.52	0.30
50.00	5.41	4.47	3.94	2.60	1.57	0.33
60.00	4.80	3.73	3.70	2.41	1.43	0.50
100.00	3.03	2.66	2.17	1.78	1.20	0.63
150.00	2.02	1.89	1.73	1.29	0.99	0.70

Note: The increase is relative to the minimum average and maximum mean squared error respectively and expressed as a percentage.

this is not the case, then one first has to decide for the maximum or the average MSE perspective. The maximum MSE perspective matches well with frequentist principles, whereas the average MSE perspective matches better with Bayesian principles. After this choice has been made, one can select the best estimator using **Tables 2** or **4**. In case constraints do not allow such an individualized choice, I recommend using the positive-part Ezekiel estimator. This choice is especially defensible in situations where the sample size *N* is large compared to the number of predictors *p*, as here the difference between estimators is small. Thus, choosing the default estimator, which most readers know, is a sound strategy.

If it is not possible to determine what is more important – unbiasedness or minimization of MSE – I recommend using the unbiased exact Olkin-Pratt estimator. There are three reasons for this. The first one is consistency. The uniformly lowest MSE unbiased perspective is the standard in regression analysis. In particular, the rationale given for using the ordinary least squared regression coefficients is that they are uniformly lowest MSE unbiased (Cohen, Cohen, West, & Aiken, 2003, p. 124). Second, psychology is traditionally primarily concerned with explanatory modeling (Yarkoni & Westfall, 2017), and in explanatory modeling, unbiasedness should be preferred over lower MSE (Shmueli, 2010). Third, and in relation to this, the fact that an estimator is unbiased guarantees that when estimates for the same property

of interest from multiple studies are aggregated in a meta-analysis, this aggregation eventually leads to the true value, which is not the case when using a biased estimator, even if it has lower MSE.

On the surface, these recommendations conflict with previous recommendations based on simulation studies with similar designs. Yin and Fan (2001) recommended using the Pratt estimator and Shieh (2008) the positive-part version of the Pratt estimator. The conflict with Yin and Fan (2001) is quickly resolved. They considered only the Pratt and the Claudy approximation of the Olkin-Pratt estimator but not the more elaborate versions. Additionally, they used bias as the only metric for comparison. Thus, were they to repeat their study with the estimator included in this comparison, they should conclude that the exact Olkin-Pratt estimator should always be used.

The conflict with Shieh (2008) can also be resolved. While Shieh (2008) considered bias and MSE, he eventually based his conclusions almost exclusively on bias. Additionally, he considered computational complexity and the estimator not returning impossible negative values as factors. Balancing all these factors lead to the positive-part Pratt estimator. Using this same balance of factors, I would expect that the results presented here would not change the recommendation by Shieh (2008).

While the rationale presented in Shieh (2008) is sound, the rationale given here has several advantages. First, computational complexity becomes an irrelevant factor due to the R package provided, which can compute all estimators within milliseconds. Second, while ignoring estimators that do not return impossible negative values is intuitively appealing and beneficial from a pure MSE-based perspective, it is detrimental from the more prevalent unbiasedness perspective. As I already mentioned, the fact that an estimator is unbiased guarantees that when estimates for the same property of interest are obtained based on multiple studies, then the average of these estimates converges to the true value. Consequently, from this perspective, returning impossible values on one sample is less detrimental than converging on the wrong value when averaging across many samples (see, Okada, 2017, for a paper-length elaboration of this argument).

Limitations and Future Work

All results presented here rely on the assumption of a multivariate normal distribution of the predictor variables. This limitation is shared with all previous comparisons. As such, repeating this study with different distributions for the predictor variables to investigate the robustness of the results with regard to this assumption is recommended for future work.

Assessing bias and MSE through a simulation study has several disadvantages. First, the values are estimated and not computed exactly. I mitigated this issue by employing a much larger number of replications than previous comparisons (100,000) and hypothesis tests to account for the remaining small uncertainty of the estimates. I

did not use more precise alternatives such as analytical derivations or numerical methods (Shieh, 2008) because both approaches would not allow a direct assessment of the provided R package.

A second disadvantage is that the conclusions from a simulation study often do not generalize beyond the considered design. I diminished this issue by augmenting the results of the simulation study with theoretical results. As such, the central finding that the Olkin-Pratt estimator is uniformly minimum MSE unbiased generalizes beyond the design considered. Whether the outcome that overall, the Ezekiel estimator performs best in terms of MSE also generalizes to other designs remains to be investigated. Also, selecting the MSE optimal estimator is only possible if the parameters of a data set (sample size, number of predictors) lie within the range considered here. For this reason, I carefully selected the parameter ranges such that they cover the majority of parameter values reported in psychology. Nevertheless, providing a table for all relevant parameter combinations is impossible. Instead, I advise researchers to run their own small simulation studies to determine the MSE optimal estimator for their particular situation. To facilitate this, I share the code used for running the simulation study as supplementary material <https://doi.org/10.24433/CO.8023088.v3>.

Conclusion

In conclusion, I recommend using the exact Olkin-Pratt estimator by default. However, if the researcher is confident that minimizing MSE is more critical than unbiasedness, then a different estimator should be used. In this case, I recommend an individualized choice based on the strategy described at the beginning of this discussion, and if this is not feasible or the sample size N is large compared to the number of predictors p , the positive-part version of the Ezekiel estimator.

Date Accessibility Statement

The data and code used to produce this paper and the supplementary material can be found at <https://doi.org/10.24433/CO.8023088.v3>.

Additional Files

The additional files for this article can be found as follows:

- **Supplemental Bias CSV File.** CSV File S1. DOI: <https://doi.org/10.1525/collabra.343.s1>
- **Supplemental MSE CSV File.** CSV File S2. DOI: <https://doi.org/10.1525/collabra.343.s2>
- **Supplemental Bias and MSE Tables.** Tables S3 to Tables S25. DOI: <https://doi.org/10.1525/collabra.343.s3>

Acknowledgements

Sarah Makonnen provided helpful feedback on the manuscript.

Competing Interests

The author has no competing interests to declare.

References

- Albers, C. J., Kiers, H. A. L., & van Ravenzwaaij, D.** (2018). Credible confidence: A pragmatic view on the frequentist vs Bayesian debate. *Collabra: Psychology*, 4(1), 31. DOI: <https://doi.org/10.1525/collabra.149>
- Alf, E. F., & Graf, R. G.** (2002). A new maximum likelihood estimator for the population squared multiple correlation. *Journal of Educational and Behavioral Statistics*, 27(3), 223–235. DOI: <https://doi.org/10.3102/10769986027003223>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., ... Johnson, V. E.** (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10. DOI: <https://doi.org/10.1038/s41562-017-0189-z>
- Cohen, J., Cohen, J., West, S. G., & Aiken, L. S.** (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Erlbaum.
- Fisher, R. A.** (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society, A*, 121, 654–673. DOI: <https://doi.org/10.1098/rspa.1928.0224>
- Gauss Hypergeometric Function 2F1: Specific Values.** (2019). <http://functions.wolfram.com/HypergeometricFunctions/Hypergeometric2F1/03/06/07/02/>.
- Karch, J. D., Filevich, E., Wenger, E., Lisofsky, N., Becker, M., Butler, O., ... Kühn, S.** (2019). Identifying predictors of within-person variance in MRI-based brain volume estimates. *Neuroimage*. DOI: <https://doi.org/10.1016/j.neuroimage.2019.05.030>
- Lehmann, E. L., & Casella, G.** (2003). *Theory of point estimation* (2nd ed. 1998. Corr. 4th printing 2003). New York: Springer.
- Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B.** (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills*, 112(2), 331–348. DOI: <https://doi.org/10.2466/03.11.PMS.112.2.331-348>
- NIST Digital Library of Mathematical Functions.** (2019). <http://dlmf.nist.gov/>, Release 1.0.23 of 2019-06-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, (eds.).
- Okada, K.** (2017). Negative estimate of variance-accounted-for effect size: How often it is obtained, and what happens if it is treated as zero. *Behavior Research Methods*, 49(3), 979–987. DOI: <https://doi.org/10.3758/s13428-016-0760-y>
- Olkin, I., & Pratt, J. W.** (1958). Unbiased estimation of certain correlation coefficients. *The Annals of Mathematical Statistics*, 29(1), 201–211. DOI: <https://doi.org/10.1214/aoms/1177706717>
- Pearson, J. W., Olver, S., & Porter, M. A.** (2015). Numerical Methods for the Computation of the Confluent and Gauss Hypergeometric Functions. *arXiv*: <https://arxiv.org/abs/1407.7786>
- R Core Team.** (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F.** (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement*, 21(4), 291–305. DOI: <https://doi.org/10.1177/01466216970214001>
- Shieh, G.** (2008). Improved shrinkage estimation of squared multiple correlation coefficient and squared cross-validity coefficient. *Organizational Research Methods*, 11(2), 387–407. DOI: <https://doi.org/10.1177/1094428106292901>
- Shmueli, G.** (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. DOI: <https://doi.org/10.1214/10-STS330>
- Yarkoni, T., & Westfall, J.** (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. DOI: <https://doi.org/10.1177/1745691617693393>
- Yin, P., & Fan, X.** (2001). Estimating R^2 shrinkage in multiple regression: A comparison of different analytical methods. *The Journal of Experimental Education*, 69(2), 203–224. DOI: <https://doi.org/10.1080/00220970109600656>

Appendix A

Computation of the Hypergeometric Function

The hypergeometric function for real inputs $a, b, c, z, \in \mathbb{R}$ is as an infinite sum ${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} t_k$ with addends

$$t_k = \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}.$$

$(q)_k$ denotes the rising factorial, which is defined by

$$(q)_k = \begin{cases} 1 & k = 0 \\ q(q+1) \cdots (q+k-1) & k > 0 \end{cases}$$

In general, the hypergeometric function is difficult to compute and for some inputs it is even not defined since it does not converge.

However, for the evaluation of the Olkin-Pratt estimator, the hypergeometric function only needs to be evaluated at the inputs $a = 1, b = 1, c \in \{m/2 : m \in \mathbb{N} \setminus \{0, 1\}\}$, and $z \in [0, 1] \subset \mathbb{R}$. I restrict this further to the case $c > 2$, as this guarantees convergence. This restriction is fulfilled if $N - p \geq 3$. Thus, there must be at least 4 samples more than there are predictors. For these inputs, closed-form solutions exist. For notational unclutteredness, I define the following function $F(c, z) = {}_2F_1(1, 1; c; z)$.

First, I discuss the pathological cases $z = 1$ and $z = 0$. For $z = 0$, $F(c; 0) = 0$. For $z = 1$, $F(c; 1) = \frac{(c-1)}{(c-2)}$.

Beyond those pathological cases, two cases have to be distinguished: 1) c being an integer, and 2) c not being an integer but divisible by 0.5.

For 1) c being an integer, a closed-form solution is presented in “Gauss Hypergeometric Function 2F1: Specific Values” (2019). It is:

$$F(c; z) = \frac{(c-1)z}{(z-1)^2} \left(\sum_{k=2}^{c-1} \frac{1}{c-k} \left(\frac{z-1}{z} \right)^k - \left(\frac{z-1}{z} \right)^c \log(1-z) \right)$$

For 2) c not being an integer but divisible by 0.5, the following rule presented in “NIST Digital Library of Mathematical Functions” (2019, Equation 15.5.16) can be leveraged.

$$F(c+1; z) = \frac{c {}_2F_1(0, 1; c; z) - c(1-z)F(c; z)}{(c-1)z}$$

Substituting ${}_2F_1(0, 1; c; z) = 1$, this becomes the following recursive equation

$$F(c+1; z) = \frac{c - c(1-z)F(c; z)}{(c-1)z}$$

with the base $F(\frac{1}{2}; z)$. The value of the base is known

$$F\left(\frac{1}{2}; z\right) = \frac{1}{1-z} \left(1 + \frac{\sqrt{z} \arcsin(\sqrt{z})}{\sqrt{1-z}} \right)$$

These observations lead to the following algorithm for calculating the function $F(c; z)$, which I display in the form of R code.

```
hypergeo_analytical <- function(c, z) {
  if (z == 0) {
    return(1)
  } else if (z == 1) {
    return((c - 1) / (c - 2))
  } else if (c %% 1 == 0) {
    shared <- ((z - 1) / z)
    the_sum <- 0
    for (k in seq(from = 2, to = (c - 1), by = 1)) {
      the_sum <- the_sum + shared^k / (c - k)
    }
    prefactor <- (c - 1) * z * (z - 1)^(-2)
    result <- prefactor * (the_sum - shared^c * log(1 - z))
  } else {
    cur_res <- 1 / (1 - z) * (1 + (sqrt(z) * asin(sqrt(z))) / sqrt(1 - z))
    for (i in seq(from = 1.5, to = c, by = 1)) {
      cur_c <- i - 1
      cur_res <- (cur_c - cur_c * (1 - z) * cur_res) / (z * (cur_c - 1))
    }
    result <- cur_res
  }
  return(result)
}
```

While this algorithm works in general, it returns wrong results for some inputs if implemented using floating-point arithmetic, as is commonly used in R and most other programming languages. In particular, the case $z < 0.5$ leads to wrong results. To resolve this, I relied on recent work about how to accurately compute the hypergeometric function using floating-point arithmetic (Pearson, Olver, & Porter, 2015).

In particular, I used the Taylor series approach, which is described in detail in Pearson et al. (2015, Section 4.2). The Taylor series approach works similarly to the approximation of the hypergeometric series in the OKP estimators. It also relies on approximating the hypergeometric series by only employing a finite number of addends t_k up to some threshold K . However, the crucial difference is that the threshold K is not fixed but is chosen dynamically such that the approximation error is minimal (see, Pearson et al., 2015, Section 4.2 for details).

To estimate the discrepancy between the Taylor series approach and the analytical algorithm, I used an input range for which the analytical algorithm returns correct results; in particular, $z \in (0.5, 0.99)$, and $c \in [5, 1000]$. In this range, the maximum difference between the Taylor series approach and the analytic algorithm was $< 10^{-12}$. Thus, the Taylor series approach and the analytical algorithm are virtually indistinguishable. Consequently, while the Taylor series approach is an approximation, its approximation error is so small that I consider it exact.

Appendix B

Data Generation Details

Here, I will explain how I simulated data for a given combination of sample size N , number of predictors p , and squared multiple correlation ρ^2 .

As I mentioned in the main text, the core assumption of this paper is that the predictors have a multivariate normal distribution with mean vector μ_x and covariance matrix Σ_x . As the mean vector μ_x , I used the vector of p zeros and as the covariance matrix Σ_x the $p \times p$ identity matrix. To obtain N observations for each predictor, I sampled from this distribution N times.

Generating the dependent variable such that it leads to a certain squared multiple correlation values ρ^2 is more complex. As the first step, I arbitrarily set the error variance of σ_ϵ^2 to 10. By rearranging Equation 2 to

$$\sigma_y^2 = \frac{\sigma_\epsilon^2}{1 - \rho^2},$$

the problem reduces to generating the dependent variable Y such that it has a total variance of σ_y^2 . The total variance σ_y^2 can be expressed as a function of the regression coefficients $\beta = [\beta_1, \dots, \beta_p]^T$, the error variance σ_ϵ^2 , and the covariance matrix of the predictors Σ_x :

$$\sigma_y^2 = \beta^T \Sigma_x \beta + \sigma_\epsilon^2. \tag{3}$$

There is still an infinite number of regression weights β that lead to a given total variance σ_y^2 . To solve this problem, I used the restriction that all regression weights must be equal. I denote their value with β_* . Using this restriction, and the fact that the covariance matrix Σ_x was the identity matrix, Equation 3 simplifies to

$$\sigma_y^2 = p\beta_*^2 + \sigma_\epsilon^2$$

and the value β_* for the regression coefficients is thus

$$\beta_* = \sqrt{\frac{\sigma_y^2 - \sigma_\epsilon^2}{p}}$$

With the error variance σ_ϵ^2 and the regression coefficients β , the dependent variable can be generated by applying Equation 1. As intercept β_0 , I used 100.

Appendix C

Estimators With the Lowest Mean Squared Error

Table C1: Estimator with the lowest mean squared error for each sample size N , number of predictors p , and multiple squared correlation ρ^2 combination.

N	p	ρ^2									
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10.00	2.00	E ⁺	E ⁺	E ⁺	S ⁺	R ²	R ²	R ²	R ²	R ²	C ⁺
10.00	5.00	E ⁺	E ⁺	E ⁺	E ⁺	S ⁺	W ⁺	R ²	R ²	R ²	R ²
20.00	2.00	E ⁺	E ⁺	E ⁺	R ²	R ²	R ²	R ²	R ²	R ²	C ⁺
20.00	5.00	E ⁺	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	R ²	R ²	R ²	R ²
20.00	10.00	E ⁺	E ⁺	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	W ⁺	R ²
30.00	2.00	E ⁺	E ⁺	W ⁺	R ²	R ²	R ²	R ²	R ²	R ²	C ⁺
30.00	5.00	E ⁺	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	R ²	R ²	R ²	R ²
30.00	10.00	E ⁺	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	C ⁺	C ⁺	C ⁺
40.00	2.00	E ⁺	E ⁺	W ⁺	R ²	R ²	R ²	R ²	R ²	R ²	C ⁺
40.00	5.00	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	R ²	R ²	R ²	R ²
40.00	10.00	E ⁺	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	C ⁺	C ⁺	C ⁺
50.00	2.00	E ⁺	E ⁺	ML	ML	R ²	R ²	R ²	R ²	R ²	C ⁺
50.00	5.00	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	R ²	R ²	R ²	R ²
50.00	10.00	E ⁺	E ⁺	S ⁺	W ⁺	W ⁺	W ⁺	W ⁺	C ⁺	C ⁺	C ⁺
60.00	2.00	E ⁺	E ⁺	ML	ML	R ²	R ²	R ²	R ²	C ⁺	C ⁺
60.00	5.00	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	R ²	R ²	R ²	R ²
60.00	10.00	E ⁺	E ⁺	W ⁺	W ⁺	W ⁺	W ⁺	W ⁺	C ⁺	C ⁺	C ⁺
100.00	2.00	E ⁺	S ⁺	ML	ML	R ²	R ²	R ²	R ²	C ⁺	C ⁺
100.00	5.00	E ⁺	S ⁺	ML	ML	ML	W ⁺	R ²	R ²	R ²	R ²
100.00	10.00	E ⁺	E ⁺	W ⁺	ML	W ⁺	W ⁺	C ⁺	C ⁺	C ⁺	C ⁺
150.00	2.00	E ⁺	ML	ML	ML	R ²	R ²	R ²	R ²	C ⁺	C ⁺
150.00	5.00	E ⁺	S ⁺	ML	ML	ML	W ⁺	R ²	R ²	R ²	R ²
150.00	10.00	E ⁺	S ⁺	W ⁺	ML	ML	W ⁺	C ⁺	C ⁺	C ⁺	C ⁺

Note: E⁺ = positive-part Ezekiel; S⁺ = positive-part Smith; ML = Maximum Likelihood; R² = normal R²; W⁺ = positive-part Wherry; C⁺ = positive-part Claudy.

The columns for $p^2 = 0$ and 0.01 are not shown. In this conditions, the positive-part Ezekiel estimator is always best as for the $p^2 = 0.05$ column.

Peer review comments

The author(s) of this paper chose the Open Review option, and Streamlined Review option, and all new peer review comments (but not ported comments from the prior review) can be downloaded at:

<http://doi.org/10.1525/collabra.343.pr>

How to cite this article: Karch, J. (2020). Improving on Adjusted R-Squared. *Collabra: Psychology*, 6(1): 45. DOI: <https://doi.org/10.1525/collabra.343>

Senior Editor: Don van Ravenzwaaij

Editor: Don van Ravenzwaaij

Submitted: 12 March 2020

Accepted: 01 July 2020

Published: 29 September 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



UNIVERSITY
of CALIFORNIA
PRESS

Collabra: Psychology

Collabra: Psychology is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS