

## **Incremental Criterion Prediction of Personality Facets over Factors: Obtaining Unbiased Estimates and Confidence Intervals**

Jeromy Anglim, Sharon L. Grant<sup>1</sup>

**Citation:** Anglim, J., & Grant, S. L. (2014). Incremental Criterion Prediction of Personality Facets over Factors: Obtaining Unbiased Estimates and Confidence Intervals. *Journal of Research in Personality*, 53, 148-157. <http://dx.doi.org/10.1016/j.jrp.2014.10.005>

### **Abstract**

Many researchers have argued that higher order models of personality such as the Five Factor Model are insufficient, and that facet-level analysis is required to better understand criteria such as well-being, job performance, and personality disorders. However, common methods in the extant literature used to estimate the incremental prediction of facets over factors have several shortcomings. This paper delineates these shortcomings by evaluating alternative methods using statistical theory, simulation, and an empirical example. We recommend using differences between Olkin-Pratt adjusted r-squared for factor versus facet regression models to estimate the incremental prediction of facets and present a method for obtaining confidence intervals for such estimates using double adjusted-r-squared bootstrapping. We also provide an R package that implements the proposed methods.

**Keywords:** personality, Big Five traits, facets, multiple regression, bootstrapping

### **1. Introduction**

Personality trait researchers have long been interested in how many personality traits are required to adequately capture individual differences. Hierarchical models of personality traits provide multiple levels of description, typified by the Five Factor Model in which the global factors of extraversion, neuroticism, conscientiousness,

---

<sup>1</sup> This article may not exactly replicate the authoritative document published in the journal. It is not the copy of record. The above doi links to the copy of record. Supplementary data and code is also available from the above doi.

Jeromy Anglim, School of Psychology, Deakin University; Sharon Grant, Faculty of Health, Arts and Design, Swinburne University of Technology. We thank Sue Carmen for her assistance with data collection. Correspondence concerning this article should be addressed to Jeromy Anglim, School of Psychology, Deakin University, 221 Burwood Highway, Burwood, 3125 Victoria, Australia. Email: [jeromy.anglim@deakin.edu.au](mailto:jeromy.anglim@deakin.edu.au)

agreeableness and openness each consist of six facets representing a more detailed level of personality. Despite the popularity of the Big Five, there has been substantial debate about the relative merits of factor and facet assessments of personality (Ashton, 1998; Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Ashton, Paunonen, & Lee, 2014; Christiansen & Robie, 2011; O'Neill, Paunonen, Christiansen, & Tett, 2013; Paunonen, 1998; Paunonen & Ashton, 2001; Paunonen, Haddock, Forsterling, & Keinonen, 2003; Paunonen, Rothstein, & Jackson, 1999; Salgado, Moscoso, & Berges, 2013). Additionally, comparing the predictive value of a model with 30 facet predictors to one with only five factor predictors has presented a challenge for researchers concerned with issues of over fitting. Personality researchers seeking to predict outcomes such as well-being (Siegler & Brummett, 2000), job performance (Ashton, 1998; Christiansen & Robie, 2011; Ones & Viswesvaran, 1996; Salgado et al., 2013; Tett, Steele, & Beauregard, 2003), and personality disorders (Bagby, Costa, Widiger, Ryder, & Marshall, 2005; Dyce & O'Connor, 1998) have then had to decide whether to include facets or only the Big Five factors as predictors.

Typically, incremental prediction of facets over factors has been estimated by subtracting the variance explained in a criterion by factors from that explained by facets. However, researchers have used many different estimators of variance explained, including unadjusted *r*-squared, adjusted *r*-squared, and cross-validated *r*-squared, combined with different regression procedures including direct entry (Mershon & Gorsuch, 1988) and stepwise regression (Baudin, Aluja, Rolland, & Blanch, 2011; Dyce & O'Connor, 1998; Ekehammar & Akrami, 2007; Quevedo & Abella, 2011; Schimmack, Oishi, Furr, & Funder, 2004); some studies have simply reported zero-order correlations (e.g., Rothmann & Coetzer, 2002; Siegler & Brummett, 2000). Thus, a principled selection of estimators is lacking (e.g., see critical review by O'Connor & Paunonen, 2007). Furthermore, the use of small sample sizes (e.g., Ashton et al., 1995; Mershon & Gorsuch, 1988; Schimmack et al., 2004) and incomplete facet-factor comparisons, based on the selection of subsets of either facets or factors (e.g., Ashton et al., 1995; Bagby et al., 2005; Dudley, Orvis, Lebiecki, & Cortina, 2006; Fruyt, Clercq, Wiele, & Heeringen, 2006; Hastings & O'Neill, 2009; Paunonen & Ashton, 2001; Salgado et al., 2013; Stephan, 2009), has limited the available empirical evidence regarding the overall incremental value of facets over factors. Also, existing research has not explicitly specified a population parameter of interest. Furthermore, as will be shown many existing methods that have been used for estimating incremental prediction result in biased estimates. The lack of reporting of confidence intervals further compounds these issues. Further clarity is needed about these foundational issues in order to more clearly quantify the gains that can be achieved by the inclusion of facets in predictive models. Thus, existing approaches are insufficient for researchers seeking to make

conclusions about the relative utility of facet- versus factor-level analysis in personality research.

The purpose of the current paper is (1) to identify the population parameter of interest for research on incremental prediction of facets over factors; (2) to compare methods for obtaining an estimate of this population parameter to demonstrate relative bias across methods through a series of simulations, and (3) to provide a method for reporting confidence intervals around this estimate. We also critically review the broader set of approaches that have been used to compare factor versus facet prediction of criterion variables. Based on our comparison of methods, we recommend the use of the Olkin-Pratt adjusted r-squared as an estimator, and the reporting of double-adjusted r-squared bootstrap (DAB) confidence intervals. We also review and make recommendations regarding methods for identifying which particular facets are of greatest incremental benefit. Finally, we present an R package that implements all the proposed methods.

## 2. Identifying the Parameter of Interest for Incremental Prediction Research

The present paper focuses on the scenario where factors and facets come from a hierarchical measure, such as the 30 facets and five factors from the NEO-Personality Inventory (Costa & McCrae, 1995). Notably however, the method presented in this paper could readily be extended to scenarios where factors and facets are not derived from a hierarchical measure, such as when additional facet predictors are included. We also note that various other questions can meaningfully be asked, such as how predictive validity of specific facets compares with specific factors, and whether equivalent numbers of facet predictors explain more variance than equivalent numbers of factors. However, the focus of this paper is to evaluate the relative predictive utility of a full set of facets versus factors in relation to the Five Factor Model, an issue that is important for personality research given the popularity of factor-level measures (Judge, Klinger, Simon, & Yang, 2008).

We contend that researchers interested in incremental prediction of facets over factors should focus on the change in population variance explained from a regression with facets versus a regression with factors as predictors. We can denote this difference as  $\Delta\rho^2$  (i.e., delta-rho-squared), where  $\Delta\rho^2 = \rho_{(\text{facets})}^2 - \rho_{(\text{factors})}^2$ , and where  $\rho_{(\text{facets})}^2$  and  $\rho_{(\text{factors})}^2$  correspond to population variance explained for models with facets and factors as predictors respectively. Note that factors are not included as predictors in the facet regression equation when factors are a weighted composite of facets as per hierarchical measures of personality such as the NEO-PI. We also note as advocated by Ozer (1985) that multiple rho (i.e.,  $\Delta|\rho| = |\rho_{(\text{facets})}| - |\rho_{(\text{factors})}|$ ) provides a legitimate alternative metric.

As a minor point, we advocate the use of  $\rho^2$  and  $\Delta\rho^2$  implied by random-score rather than fixed-score regression models sometimes referred to as random-x and fixed-x assumptions (for further discussion, see Fox, 2002; Yin & Fan, 2001). In a fixed-score regression, it is assumed that the values of predictors are fixed across studies. In a random-score regression it is assumed that the predictors are to be sampled from an underlying population. Given that the aim is to draw inferences about the full population of personality data, the random-score regression model is more appropriate.

It is also important to note that  $\rho^2$ , which is the variance explained in the population using the population regression equation, differs from  $\rho_c^2$  (i.e., cross-validated rho-square) which is the variance explained by the regression equation obtained in the sample when applied to the population (Yin & Fan, 2001). Specifically,  $\rho^2$  is relevant to understanding true theoretical relationships, whereas  $\rho_c^2$  is relevant where the aim is to apply a sample estimated regression equation to a practical prediction context.

We note that when facets come from a hierarchical measure where factors are defined as a weighted composite of facets, a regression with facets will always explain as much variance as or more variance than factors ( $\Delta\rho^2 \geq 0$ ). Thus, for a given criterion, once a regression approach is adopted, the question is not *whether* facets explain more variance than factors but rather *how much more* variance they explain. So, an important substantive question for personality researchers is whether the amount of incremental prediction for a given criterion justifies the increased complexity associated with the increased number of predictors.

### 3. Selecting an Estimator of Incremental Variance Explained

#### 3.1. Description of Estimators

We now review the different methods that have been used to estimate the incremental population variance explained by a regression with factors as predictors versus one with facets as predictors, denoted  $\Delta\rho^2$ . In general, estimates of  $\Delta\rho^2$  are obtained by first obtaining estimates of  $\rho^2$  for facets and for factors, and then subtracting one from the other:  $\Delta\hat{\rho}^2 = \hat{\rho}_{\text{facets}}^2 - \hat{\rho}_{\text{factors}}^2$ , i.e., where the hats indicate estimates of corresponding population parameters. Three major classes of estimators of  $\rho^2$  are (a) unadjusted r-squared (i.e.,  $R^2$ ), (b) adjusted r-squared (i.e.,  $R_{\text{adj}}^2$ ), and (c) cross-validated r-squared (i.e.,  $R_c^2$ ). Typically, but not always, use of adjusted r-squared has been combined with direct entry of all factors or all facets as predictors, and unadjusted r-squared has been

used with stepwise entry of factors or facets, whereas cross-validated r-squared has rarely been used in the facet–factor comparison literature.

Unadjusted r-squared is the variance explained in the sample data by the sample estimated regression equation. Cross-validated r-squared represents a broad class of techniques that attempt to estimate  $\rho_c^2$ , i.e., the population prediction using the sample regression equation. Adjusted r-squared shrinks unadjusted r-squared. The shrinkage is greater when sample sizes are smaller and the number of predictors is greater. Adjusted r-squared is designed to provide an unbiased estimate of  $\rho^2$ . There are several adjusted r-squared formulas (for a review see Raju, Bilgic, Edwards, & Fler, 1997). The Ezekiel and Fox (1959) formula is commonly used in statistical packages, where

$$R_{adj(E)}^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} \quad (1)$$

and where  $n$  is the sample size and  $p$  is the number of predictors. Adjusted r-squared formulas differ based on whether they are designed to estimate fixed-x  $\rho^2$  or random-x  $\rho^2$ . In particular, the standard Ezekiel formula shown above is an estimator of fixed-x  $\rho^2$  whereas the Olkin-Pratt (1958)  $R_{adj(OP)}^2$  formula and several other approximations are designed to estimate random-x  $\rho^2$ . Specifically, the Olkin-Pratt formula is

$$R_{adj(OP)}^2 = 1 - (1 - R^2) \frac{n-3}{n-p-1} F \left[ 1, 1; \frac{n-p+1}{2}; (1 - R^2) \right] \quad (2)$$

where  $F$  is the hypergeometric function. As discussed earlier, given that personality research samples the predictor values from a population, researchers should be making the random-x assumption. Thus, we recommend the use of the adjusted r-squared using the Olkin-Pratt formula to estimate  $\Delta\rho^2$ .

### 3.2. Evaluating Estimators of Incremental Variance Explained

We now examine the theoretical properties of the estimators discussed above. First, if  $\rho_{(\text{facets})}^2$  and  $\rho_{(\text{factors})}^2$  are estimated in an unbiased way, then  $\Delta\hat{\rho}^2 = \hat{\rho}_{\text{facets}}^2 - \hat{\rho}_{\text{factors}}^2$  will be an unbiased estimate of  $\Delta\rho^2$ . Likewise, if the bias is equal for both factors and facets [i.e.,  $\rho_{\text{facets}}^2 = E(\hat{\rho}_{\text{facets}}^2) + c$  and  $\rho_{\text{factors}}^2 = E(\hat{\rho}_{\text{factors}}^2) + c$ , where  $c$  is the bias, and  $E$  is the expectation operator], then  $\Delta\hat{\rho}^2$  will also be unbiased. Thus, we first focus on obtaining unbiased estimates of  $\rho_{(\text{facets})}^2$  and  $\rho_{(\text{factors})}^2$ .

To start let us assume that the complete set of either facet or factor predictors have been entered into the regression model. In this case, unadjusted r-squared is a positively biased estimate of  $\rho^2$  where the bias is larger for smaller sample sizes and more predictors (Barten,

1962). Most adjusted r-squared formulas take unadjusted r-squared, sample size, and the number of predictors and aim to provide an unbiased estimate of  $\rho^2$ . As sample size goes to infinity, the shrinkage factor applied by adjusted r-squared formulas approaches one and thus there is no difference between adjusted and unadjusted r-squared. Furthermore, in the limit, the sample regression equation approaches the population regression equation, and the sample distribution of predictor data approaches the population distribution of predictor data. While with an infinite sample size  $R_{\text{adj}}^2 = R^2 = \rho^2$ , actual research employs finite sample sizes. A few studies have estimated  $\Delta\rho^2$  using  $R_{(\text{facets})}^2 - R_{(\text{factors})}^2$ . Such an estimator is positively biased because there are more predictors in the facet regression equation than in the factor regression equation.

If adjusted r-squared is an unbiased estimate of  $\rho^2$  then  $R_{\text{adj}(\text{facets})}^2 - R_{\text{adj}(\text{factors})}^2$  is an unbiased estimate of  $\Delta\rho^2$ . In general, random-x estimates of  $\rho^2$  yield slightly larger estimates, but the difference is small if sample size is moderately large, e.g.,  $n > 200$  (for some specific comparisons, see Leach & Henson, 2007), and the difference disappears asymptotically. This contrasts with much of the extant literature, which has frequently used alternative data driven approaches for predictor selection, including stepwise regression and *best facet* selection i.e., using a set of predictors composed of the one facet from each factor that correlates most with the criterion (e.g., Hastings & O'Neill, 2009). Because of the popularity of such data driven approaches, we review the properties of such estimators in the Appendix. While we discourage the use of such data driven approaches for estimating  $\Delta\rho^2$ , an understanding of their estimation properties is important for appraising the extant literature. Having evaluated methods for deriving unbiased estimators of  $\Delta\rho^2$ , we now focus on quantifying the uncertainty in these estimates.

#### 4. Confidence Intervals on Incremental Variance Explained

There has been a movement in the reporting of psychological results that advocates the reporting of confidence intervals on effect sizes. Despite this recommendation, we were unable to find a single study in the personality factor versus facet comparison literature that has reported a confidence interval for  $\Delta\rho^2$ . This is presumably due to the lack of integration in standard statistical software for confidence intervals for  $\rho^2$ , let alone  $\Delta\rho^2$ . Thus, a major contribution of this paper is the provision of an R package to enable personality researchers to calculate confidence intervals for estimates of  $\Delta\rho^2$  in assessing the incremental prediction of facets over factors.

There are several approaches for obtaining confidence intervals on  $\Delta\rho^2$ . Alf Jr and Graf (1999) outlined how to obtain asymptotic confidence intervals. Algina, Keselman, and Penfield (2007) applied a percentile bootstrap on unadjusted r-squared change, and dismissed a simple percentile bootstrap based on sample adjusted r-squared change. However, it is clear that  $R^2_{(\text{facets})} - R^2_{(\text{factors})}$  is a biased estimator of  $\Delta\rho^2$  and as a consequence any percentile bootstrap based on this would also be biased. While this bias is minor when comparing regressions with a similar number of predictors, the bias is problematic when comparing a regression with 30 facet predictors to one with five factor predictors. We propose a Double Adjusted r-squared Bootstrap (DAB) method that serves to remove this bias. The procedure is as follows:

1. From the original dataset of  $n$  cases, obtain  $K$  bootstrapped samples each of size  $n$  sampled with replacement from the original dataset.
2. For each bootstrapped sample:
  - a. Calculate  $R^2_{\text{adj}}$  for the factor and facet linear regressions. Assuming the predictors are a random sample, use the Olkin-Pratt adjusted r-squared formula.
  - b. Reapply the adjusted r-squared formula using the  $R^2_{\text{adj}}$  value for factors and facets from step 2a, using  $R^2_{\text{adj}}$  in place of  $R^2$  in the adjusted r-squared formula; let us denote this new double-adjusted r-squared value  $\tilde{R}^2_{\text{adj}}$ .
  - c. Estimate  $\Delta\rho^2$  for the bootstrapped dataset as  $\tilde{R}^2_{\text{adj}(\text{facets})} - \tilde{R}^2_{\text{adj}(\text{factors})}$ .
3. Take the .025 and .975 quantiles of the bootstrapped estimates of  $\Delta\rho^2$  to obtain a 95% confidence interval and the standard deviation as the standard error.

The first application of the adjusted r-squared formula removes the bias that results from bootstrapping assuming that the sample dataset is the population. The second application of adjusted r-squared provides the unbiased estimate as per a standard non-bootstrapped sample. We recommend using at least  $K=10,000$  bootstrapped samples for publication results, unless computational time is prohibitive. For example, at the time of writing, using a modern desktop computer, the bootstrap procedure with 337 cases, 30 facets, 5 factors and  $K=10,000$  took less than a minute to run. More importantly, under these conditions, the standard deviation in .025 and .975 confidence interval values associated with repeated runs of the bootstrap simulation was approximately .0005. This is quite

reasonable given that such confidence intervals would typically be reported to three decimal places.

### 5. The Current Study: A Demonstration of Relative Bias across Estimation Methods

The preceding sections have critically evaluated estimators of  $\Delta\rho^2$  that have been used in the literature. The analyses that follow apply the different estimators to a sample dataset and demonstrate the proposed procedure for obtaining confidence intervals. We then present a simulation study to assess bias and standards error of common estimators of  $\Delta\rho^2$  for different sample sizes and datasets. Specifically, it was expected that Olkin-Pratt adjusted r-squared would provide the best estimator of  $\Delta\rho^2$ . Finally, we show how semi-partial correlations can be applied to identify facets that provide incremental prediction over and above personality factors. All data and data analysis code in R used to run the illustration and the simulation is provided as an online supplement.

### 6. Illustration of Recommended Method

The current section compares different estimators of incremental variance explained and applies the Double Adjusted r-squared Bootstrap (DAB) method to calculate confidence intervals, to a sample dataset. The data comes from a study (Anglim & Grant, in press) completed by 337 adults recruited from two Australian universities (76% female; mean age of 24.4, SD = 8.8). The 300 item version of the International Personality Item Pool modeled on the NEO-PI (Goldberg, 1999) was used to measure the five factors and 30 facets. All items were completed on a 1 to 5 scale from 1=very inaccurate to 5=very accurate. Mean Cronbach's alpha reliability was .806 for factors and .803 for facets. The criterion, satisfaction with life, was measured using Diener's (1985) 5-item Satisfaction with Life Scale. Responses were on a 7-point scale from 1=strongly disagree to 7=strongly agree. After item reversal, all scale scores were computed as the mean of specified items.

We included both our recommended Olkin-Pratt  $R^2_{adj}$  estimator along with the more commonly used Ezekiel  $R^2_{adj}$  estimator. For comparison purposes we also included unadjusted  $R^2$ , two stepwise regression methods (standard and penalized), best facets  $R^2$ , and cross-validated  $R^2$  (See the Appendix for more information on stepwise, best facets, and cross-validated estimators). Unadjusted  $R^2$  used the unadjusted sample r-squared for factors and facets. Standard stepwise involved performing stepwise regression with a .05 p-value for predictor entry and then estimating  $\Delta\hat{\rho}^2 = R^2_{(facets)} - R^2_{(factors)}$ . While penalized stepwise used the same .05 p-value for factor predictor entry, it differed



from standard stepwise in that it used .0083 p-value for facet predictor entry (i.e.,  $.05 / (5 / 30) = .0083$ , where 5 corresponds to the number of factors, 30 corresponds to the number of facets and .05 is the p-value for predictor entry in the factor multiple regression). Penalized stepwise was included as an estimator because it has been presented as a way of compensating for the different number of factor and facet predictors (Paunonen, 1998). Best facet correlates used unadjusted r-squared of all factors for the factor regression and the facet from each factor with the largest absolute correlation with the criterion for the facet regression (i.e., one facet from each factor was retained). Cross-validated  $R^2$  used k-fold cross-validation.

Using the recommended Olkin-Pratt formula, the estimated incremental variance explained by facets over factors in the criterion was .117 with a 95% confidence interval of .056 to .182 (10,000 replications, using the DAB method). Thus, the ratio of facet to factor variance explained was estimated to be 1.291.

Table 1 shows the estimated variance explained in the criterion (satisfaction with life) by factors and facets using the different estimators. For this dataset, there were minimal differences between the two adjusted r-squared formulas consistent with an expectation that as sample size gets moderately large, such differences become negligible. Stepwise regression estimates were also fairly similar to adjusted r-squared estimates. Standard stepwise estimates were slightly larger and penalized stepwise estimates were slightly smaller. As expected, unadjusted r-squared was much larger than adjusted r-squared estimates. This difference was much more pronounced for the facet regression due to the increased bias due to the greater number of facet predictors. This greater bias for facet than for factor regressions translates into a much larger estimate of incremental variance explained.

In contrast, cross-validated  $R^2$  was lower than adjusted r-squared estimates for factors and much lower for facets, which resulted in a much lower value for incremental variance explained. This is consistent with expectations around cross-validation, where more predictors increase the failure of the sample regression equation to match the prediction of the population regression equation. Best facets  $R^2$  had a lower estimate of incremental variance explained. This is consistent with the expectation that as sample sizes get larger, the facet regression equation will underestimate the true regression equation, especially in the absence of an adjustment to the r-squared from the factor regression equation.

**Table 1**

*Comparison of Estimators of Population Variance Explained in Satisfaction with Life by Factors and Facets of Personality in Sample Data*

Estimator	$\hat{\rho}_{(\text{factors})}^2$	$\hat{\rho}_{(\text{facets})}^2$	$\Delta\hat{\rho}^2$
-----------	-----------------------------------	----------------------------------	----------------------

Olkin-Pratt $R^2_{adj}$	.401	.518	.117
Ezekiel $R^2_{adj}$	.399	.516	.117
Unadjusted $R^2$	.408	.559	.151
Standard Stepwise $R^2$	.408	.530	.121
Penalized Stepwise $R^2$	.408	.521	.113
Best Facets $R^2$	.408	.516	.108
Cross-validated $R^2$	.386	.460	.073

Note.  $n = 337$

Overall, these results reinforce our earlier arguments regarding expected differences between estimators and support the use of adjusted r-squared methods. The results also highlight the value of obtaining confidence intervals in order to quantify uncertainty in estimates of incremental variance explained. Interestingly, even with a relatively large sample of over 300 participants, the confidence interval was still relatively large. This is not a critique of the method for deriving the confidence intervals. As will be shown by the standard errors in the simulation presented below, this merely reflects the uncertainty inherent in estimating incremental variance explained. In particular, this highlights the importance of obtaining relatively large sample sizes, when the aim is to quantify incremental variance explained.

## 7. Simulation Comparing Estimators

In order to evaluate the properties of various estimators of  $\Delta\rho^2$  a simulation study was performed. The study assessed the bias and standard error of six estimators using two data generating mechanisms and four sample sizes. The six estimators were Olkin-Pratt adjusted r-squared, Ezekiel Adjusted r-squared, unadjusted r-squared, standard stepwise, penalized stepwise, and best facets as described in the previous data analysis section. The two data generating mechanisms labeled *no effect* and *real effect* correspond to  $\Delta\rho^2$  of .00, and .151 respectively. These two effect sizes reflect two contrasting scenarios for plausible effect sizes, one involving no effect, and one involving a moderately large effect. Specifically, the value of .151 was based on the effect size that would result from treating the previous sample data as the population; i.e., see the unadjusted r-squared value in Table 1.

Simulating data from the real effect generator involved sampling  $n$  observations from a multivariate normal distribution with means of zero, and covariance matrix equal to the sample correlation matrix from the above empirical study for the 5 factors, and the 30 facets, and the criterion (i.e., satisfaction with life). Simulating from the no effect generator was similar to the real effect generator except that only the factor and facets were sampled from the multivariate normal distribution. The criterion equaled the sum of the predicted score from the factor

regression equation and an independent normally distributed variable with a mean of zero and a standard deviation equal to the square root of the error variance in the factor regression. Thus, the real effect generator treated the covariance matrix for sample data as the population. The no effect generator was the same as real data except that the criterion only related to the factors as per the real data, but had no incremental prediction from facets.

The four sample sizes studied were 50, 100, 200, and 1,000. These sample sizes correspond roughly to our expectations of completely inadequate, inadequate, adequate, and very adequate sample sizes, respectively. Many published studies comparing personality factor and facet prediction have been between 100 and 200 (e.g., Ekehammar & Akrami, 2007; Paunonen & Ashton, 2001; Schimmack et al., 2004). A few studies have been between 50 and 100 (e.g., Mershon & Gorsuch, 1988; Paunonen, 1998; Tett et al., 2003). To perform the simulation 10,000 samples were taken for each combination of sample size and data generator. For each of these samples, an estimate of  $\Delta\rho^2$  was obtained for each of the estimators. For each combination of estimator, sample size, and data generator, estimates of bias and standard error were computed. To estimate bias and the standard error, we first calculate the mean of the simulated estimate as

$$\widehat{\Delta\rho^2} = \frac{1}{B} \sum_{i=1}^B \Delta\hat{\rho}_i^2 \quad (3)$$

where  $i=1, \dots, B$  indexes the  $i$ th simulated dataset. The estimated bias is the mean of the simulated estimates minus the true value of the estimand:

$$\widehat{\text{Bias}} = \widehat{\Delta\rho^2} - \Delta\rho^2. \quad (4)$$

Estimated standard error of the estimates is the standard deviation of the estimates:

$$\widehat{\text{SE}} = \sqrt{\frac{1}{B-1} \sum_{i=1}^B \left( \Delta\hat{\rho}_i^2 - \widehat{\Delta\rho^2} \right)^2}. \quad (5)$$

Table 2 and Table 3 report the bias and standard error for the different estimators and simulation conditions. Overall, results are consistent with the results presented in the previous section. In particular, results supported the use of both Olkin-Pratt and Ezekiel adjusted r-squared estimators, which had almost no appreciable bias under the data generating processes and sample sizes studied. Likewise, standard errors for these two adjusted r-squared estimators were similar to other estimators for sample sizes over 100. Ultimately, obtaining reasonable standard errors requires relatively large sample sizes.

**Table 2**

*Simulation Estimates of Bias in Estimating  $\Delta\rho^2$  by Estimator, Data Generating Mechanisms and Sample Size*

Estimator	n=50	n=100	n=200	n=1000
No-effect: $\Delta\rho^2 = 0$				
Olkin-Pratt $R^2_{adj}$	.000	.000	.000	.000
Ezekiel $R^2_{adj}$	.001	.000	.000	.000
Unadjusted $R^2$	.309	.152	.075	.015
Standard Stepwise $R^2$	.072	.035	.017	.003
Penalized Stepwise $R^2$	.031	.004	-.003	-.003
Best Facets $R^2$	.013	-.018	-.037	-.054
Real-effect: $\Delta\rho^2 = .151$				
Olkin-Pratt $R^2_{adj}$	-.002	.001	.000	.000
Ezekiel $R^2_{adj}$	-.002	.001	.000	.000
Unadjusted $R^2$	.214	.106	.052	.010
Standard Stepwise $R^2$	.020	.004	-.001	-.001
Penalized Stepwise $R^2$	-.013	-.024	-.023	-.010
Best Facets $R^2$	-.040	-.042	-.043	-.043

As implied by theory, stepwise, adjusted r-squared, and unadjusted r-squared all appeared to be converging to unbiased estimates as sample size increased i.e., the standard error became smaller and bias approached zero. However, unadjusted r-squared displayed very large bias even for samples sizes of 200, and appreciable bias even with sample sizes of 1,000. Stepwise regression methods, although not as good as adjusted r-squared, converged more rapidly and often showed only small amounts of bias. Using the penalized .0083 p-value for facet entry, rather than the same .05 p-value, resulted in reduced estimates of  $\Delta\rho^2$  which under several conditions led to negative bias and underestimation.

**Table 3**

*Simulation Estimates of Standard Error in Estimating  $\Delta\rho^2$  by Estimator, Data Generating Mechanisms and Sample Size*

Estimator	n=50	n=100	n=200	n=1000
No-effect: $\Delta\rho^2 = 0$				
Olkin-Pratt $R^2_{adj}$	.150	.053	.023	.004
Ezekiel $R^2_{adj}$	.148	.053	.023	.004
Unadjusted $R^2$	.082	.042	.021	.004

Standard Stepwise $R^2$	.078	.041	.022	.005
Penalized Stepwise $R^2$	.079	.048	.025	.007
Best Facets $R^2$	.053	.035	.025	.011
Real-effect: $\Delta\rho^2 = .151$				
Olkin-Pratt $R^2_{adj}$	.135	.066	.041	.016
Ezekiel $R^2_{adj}$	.135	.066	.041	.016
Unadjusted $R^2$	.089	.057	.038	.016
Standard Stepwise $R^2$	.094	.061	.041	.017
Penalized Stepwise $R^2$	.091	.059	.039	.018
Best Facets $R^2$	.065	.047	.033	.014

In contrast, the best facet method generally under estimated the incremental variance explained by facets. As sample sizes increased, the best facet procedure was less able to capitalize on additional prediction by facets due to chance. Thus, when facets had no true incremental prediction, the negative bias was amplified as sample sizes increased. When there was a real prediction of facets, the processes of increased chance prediction and refining actual estimates seemed to cancel out.

Examination of standard errors in Table 3 provides useful guidance regarding sample sizes required to achieve a given level of precision. Standard errors were larger when  $\Delta\rho^2$  was larger, although researchers are likely to be more tolerant of larger standard errors when the effect size is larger. That said, even a sample size of 200 yields moderately large uncertainty. For example, for the real effect generator, approximately 95 percent of sample estimates of  $\Delta\rho^2$  would range between .071 and .231, calculated as the population parameter plus or minus 1.96 times the standard error:  $.151 \pm (1.96 \times 0.041)$ . Thus, this supports recommendations of having sample sizes of at least 200 when precision in estimating incremental variance explained is a goal.

### 8. Identifying the Importance of Specific Personality Facets

We now make a few brief recommendations for researchers aiming to identify the role and importance of specific personality facets. If factors are taken as primary, a meaningful incremental prediction of facets over factors is necessary to further examine the facet-criterion relationship. However, where such incremental prediction is present, this raises the issue of how best to describe the pattern of relationships. Several techniques have been adopted in the literature. Researchers commonly report the zero-order correlations between facets and criterion (e.g., Rothmann & Coetzer, 2002; Siegler & Brummet, 2000). Also, some researchers report the facets that are retained in a stepwise regression (e.g., Quevedo & Abella, 2011).

Both methods have limitations. In stepwise regression, predictor inclusion includes a random component. In particular, where predictors are highly correlated which may be the case with facets, the inclusion of one facet may be particularly arbitrary. The focus is on whether the variable is a predictor rather than the strength of the relationship. Included facets take on the prediction of any variance shared with another predictor. It also does not focus on which predictors are incremental predictors over and above the Big Five factors. Focusing on zero-order facet-criterion correlations alone leads to a representation of relationships that lacks parsimony. Parsimony is a major benefit of using the Big Five factors. Thus, it is useful to see how facets provide incremental prediction over and above the factors.

A particularly useful approach that has sometimes been adopted (e.g., Fruyt et al., 2006) is to examine semi-partial correlations between facets and criterion. This first assumes that an overall understanding of factor-level prediction has been obtained by examining relationship between factors and the criterion using the zero-order correlations or standardized coefficients from a regression model with just factors. Then with that context, it can be particularly informative to compare zero-order correlations with semi-partial correlations between each facet and the criterion controlling for progressively more variables: (a) the focal factor that the facet belongs to, (b) all factors, and (c) all other facets. In particular, the semi-partial correlation controlling for all factors quantifies the degree of incremental prediction provided by the facet over the factors. Large sample sizes, a more stringent alpha (e.g.,  $p < .001$ ), and a focus on effect sizes assist in interpreting such correlations with the aim of flagging the facets contributing meaningfully to incremental prediction.

We now apply this approach to the previously presented satisfaction with life data. The overall prediction of the criterion indicated both in terms of correlations and standardized coefficients from a factor regression indicated that neuroticism ( $r = -.57$ ;  $\beta = -.37$ ) and extraversion ( $r = .51$ ;  $\beta = .31$ ) were the main factor predictors with much smaller values for conscientiousness ( $r = .35$ ;  $\beta = .11$ ), agreeableness ( $r = .11$ ;  $\beta = -.01$ ) and openness ( $r = .13$ ;  $\beta = .02$ ). Table 4 then illustrates the approach of examining facet semi-partial correlations: starting from the zero-order column, each subsequent column to the right partials out additional variables. The mean absolute correlations were .275 for zero-order, .162 controlling for the focal factor, .065 controlling for the Big Five factors, and .041 controlling for all other facets. Importantly, the zero-order correlations, which are often presented in isolation, show a large number of large correlations. In comparison, when controlling for the Big 5, a more parsimonious representation is provided whereby it is mostly depression and cheerfulness that remain important incremental predictors, as would be predicted based on the theoretical links between these facets and the criterion.

Table 4 also highlights the problems with focusing on just a subset of the Big Five factors. There are several large semi-partial correlations when controlling only for the focal factor that become small when controlling for all the Big Five factors. For example, trust goes from .36 when controlling for agreeableness only to .05 when controlling for all Big Five factors.

**Table 4**

*Zero-order and Semi-Partial Correlations of Personality Facets with Satisfaction with Life*

Focal Factor Facet	Correlation			
	Zero order	Semi-partial focal factor	Semi-partial all 5 factors	Semi-partial all other facets
Neuroticism				
Anxiety	<b>-.47</b>	.04	.02	.00
Anger	<b>-.31</b>	<b>.16</b>	.05	.01
Depression	<b>-.67</b>	<b>-.35</b>	<b>-.28</b>	<b>-.22</b>
Self-consciousness	<b>-.46</b>	-.05	<b>.14</b>	.02
Immoderation	-.16	<b>.17</b>	.09	.00
Vulnerability	<b>-.49</b>	-.01	.00	-.03
Extraversion				
Friendliness	<b>.44</b>	.04	-.03	-.05
Gregariousness	<b>.34</b>	-.14	-.05	.03
Assertiveness	<b>.39</b>	-.01	-.08	-.02
Activity level	<b>.33</b>	.09	-.02	-.02
Excitement seeking	<b>.21</b>	<b>-.17</b>	-.03	.02
Cheerfulness	<b>.57</b>	<b>.27</b>	<b>.26</b>	<b>.18</b>
Openness				
Imagination	.06	-.05	.07	.00
Artistic interests	.06	-.02	-.01	.01
Emotionality	.01	-.10	.04	-.01
Adventurousness	<b>.27</b>	<b>.25</b>	-.01	-.02
Intellect	<b>.18</b>	.13	-.06	-.03
Liberalism	-.10	<b>-.20</b>	-.03	-.03
Agreeableness				
Trust	<b>.35</b>	<b>.36</b>	.05	-.02
Morality	.05	-.07	-.02	-.05
Altruism	<b>.26</b>	<b>.27</b>	-.01	-.01
Cooperation	.07	-.02	.03	.03
Modesty	<b>-.32</b>	<b>-.45</b>	-.11	-.02
Sympathy	.09	.00	.07	.04
Conscientiousness				
Self-efficacy	<b>.49</b>	<b>.34</b>	.02	.04
Orderliness	.07	<b>-.28</b>	-.09	-.08
Dutifulness	.14	-.16	-.09	-.05
Achievement striving	<b>.42</b>	<b>.23</b>	.13	.10
Self-discipline	<b>.41</b>	<b>.21</b>	.05	.04
Cautiousness	.06	<b>-.24</b>	-.01	.04

*Note.* n = 337. Significant correlations are bolded (p < .001). Semi-partial focal factor is the semi-partial correlation between the facet and satisfaction with life adjusting the facet for the factor to which the facet



belongs. Similarly, semi-partial all 5 factors adjusts for all five factors, and semi-partial all other facets adjusts for all other facets.

## 9. General Discussion

### 9.1. Major Recommendations

The major contribution of this paper was to provide recommendations for obtaining unbiased point estimates and confidence intervals for population incremental variance explained by facets over factors and therefore answer the question of how much more variance facets explain. In particular, principled examination of estimators and simulation both supported the use of adjusted r-squared formulas. While the simulation showed minimal differences between Olkin-Pratt and Ezekiel formulas, we still encourage researchers to use the Olkin-Pratt formula given that it should yield slightly better estimates of the constituent factor and facet population variance explained. We have also presented a method for obtaining unbiased confidence intervals on population incremental variance explained particularly suited to the context where there are many more facet than factor predictors, as is the case with the Five Factor Model of personality.

The simulation reinforced the importance of obtaining moderately large sample sizes when the aim of research is to evaluate the size of incremental prediction by facets. The standard errors in the simulation provide guidance on the precision to expect under some common scenarios. While running customized simulations under various conditions provides a more accurate description of expected precision, rules of thumb can also be helpful for researchers. As a rule of thumb, we see 200 participants as the minimum required for reasonable standard errors, but the more participants the better. Results based on sample sizes less than 100 should be interpreted with caution. In addition, reporting of confidence intervals, which will be very large for small samples sizes (e.g.,  $n < 120$ ), should encourage collection of larger samples.

A major obstacle to the reporting of confidence intervals on incremental variance explained has been the lack of software that readily implements such analyses. To assist researchers in implementing the proposed methods, we have created an R package called "personalityfacets" (available from <https://github.com/jeromyanglim/personalityfacets>). The package provides estimates of  $\Delta\rho^2$  and  $\rho^2$  using Olkin-Pratt adjusted r-squared. It implements the proposed DAB method for obtaining confidence intervals on estimates of  $\Delta\rho^2$ . It also has a range of functions useful for analyzing personality facets data. This includes functions to calculate facet-criterion semi-partial correlations and associated significance tests.

### 9.2. Implications for Related Research Questions

While the present study specifically focused on factors and facets from hierarchical measures of personality, the methods discussed could

readily be applied to other approaches to selecting facets and factors. In particular, methods could be applied to studies that include external facets in the set of facet predictors (e.g., Quevedo & Abella, 2011). They could also be applied to non-hierarchical tests or to cases where facets come from a different test to where the factors come from (e.g., Paunonen & Ashton, 2001). Note that in non-hierarchical cases, judgment is required regarding whether to include the factors in the facet regression equation.

Other studies have used only a subset of facets or subset of factors that are otherwise derived from hierarchical measures of the Five Factor Model (e.g., Salgado et al., 2013). Commonly this has involved examining incremental prediction of facets from a single factor over and above that factor (e.g., Dudley et al., 2006). While the methods we propose could be applied in such settings, we recommend that researchers avoid excluding the other Big Five factors. Where only one Big Five factor is included, incremental variance provided by facets might be due to overlap with the omitted Big Five factors. All Big Five factors and their facets should be included to accurately evaluate the need for facet-level analysis in the context of the Five Factor Model.

Other studies have compared facet versus factor level prediction using a limited set of facets often derived from judges' ratings of predictor relevance to a criterion or prior research recommendations (Ashton et al., 1995; Bagby et al., 2005; Dudley et al., 2006; Fruyt et al., 2006; Hastings & O'Neill, 2009; Paunonen & Ashton, 2001; Stephan, 2009). While such approaches are problematic for evaluating incremental prediction of the full set of facets, such research may seek to answer a research question that is distinct from that addressed by the present paper. Specifically, researchers may seek to determine whether broad (commonly operationalized as factors) or narrow (commonly operationalized as facets) traits are individually more predictive of a criterion. This may be assessed by a range of methods including comparing facet–criterion with factor–criterion correlations or comparing a criterion regression model with factor predictors to one with facet predictors where the number of facets equals the number of factors. Such research is concerned with fundamental aspects of the relationship between personality and behavior and the relative primacy of narrow versus broad traits. We note that such research raises a range of issues around facet selection, multiple comparisons, reliability, and measurement that are beyond the scope of the present paper (for further discussion, see Ashton et al., 2014).

However, in general, when evaluating incremental prediction of facets, there are problems in adopting a rational rather than an empirical approach to predictor selection. It is appropriate to be rational when forming hypotheses, but empirical when evaluating hypotheses. For instance, we can rationally form an hypothesis that facets will predict much more than factors in contexts where there is a conceptual argument;

but we should empirically evaluate whether this is true. If we exclude some facets or some factors from the regression model, then we will not get an empirical answer to our question. It is for these reasons that we propose our methods. The overall rho-squared change is an empirical answer to the question of how much facets incrementally predict over factors. The semi-partial facet-criterion correlation controlling for factors provides an answer to the question of which facets provide meaningful incremental prediction.

The bifactor modeling approach also offers an alternative and complementary latent variable perspective for understanding facet-criterion relationships (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012). The bifactor model has been proposed as a way of modeling the unique contribution of facets. Such an approach typically partitions variance in personality test items into latent facets and latent factors that are all mutually orthogonal. This changes the meaning of facets and factors compared to the standard approach discussed in this paper of using summated scales. In this sense, the bifactor modeling approach leads to a conception of factors and facets so different from the summation approach that such approaches would effectively be answering different research questions. In addition to adjusting for measurement error, the bifactor approach estimates factors as the common variance across items rather than being an equally weighted formative construct based on the facets. Current applications of the bifactor model have focused on interpreting model fits and coefficients for facets and factors predicting criterion. Nonetheless, in line with the focus of this paper, it would be possible to obtain estimates of incremental variance explained by facets in a bifactor model. For example, de Jong (1999) demonstrated a method for obtaining incremental variance explained using a structural equation modeling (SEM) approach. Nonetheless, standard SEM estimation would lead to biased estimates similar to using subtraction of unadjusted  $R^2$  from multiple regression. Thus, further work is needed in order to derive unbiased estimates of incremental variance explained of facets over factors using a bifactor modeling approach. For example, future research could explore the use of adjusted r-squared formulas applied to regressions based on the implied factor covariance matrix in a bifactor model. Similar comments to the above could also be made about various higher-order confirmatory factor analytic approaches such as those recently espoused by Ashton et al. (2014).

### **9.3. Conclusions: Reinterpreting the Literature on Factor versus Facet Comparison**

The theoretical analysis and simulations presented in this study have implications for interpreting the existing substantive literature. In particular, several researchers make the claim that facets may account for double the variance of factors in relation to well-being (e.g., Quevedo &

Abella, 2011; Stephan, 2009). In the present study,  $\hat{\rho}_{(\text{facets})}^2$  was around 0.3 times larger than  $\hat{\rho}_{(\text{factors})}^2$ . Quevedo and Abella (2011) found that facets explained 0.37 times more variance than factors, and it was only after the inclusion of predictors additional to the 30 facets from the Big 5 measure that the doubling was achieved. Many other studies with different predictors have also used methods that may have overstated the difference; this is particularly so in some of the small sample studies with selective reporting of results.

In predicting personality disorders, there has been strong reliance on using a subset of facet predictors recommended by various researchers (Bagby et al., 2005). We recommend that researchers use all 30 facet predictors instead of just a subset in order to provide an empirical assessment of the degree of incremental prediction provided by facets. Several of the studies have also had fairly small sample sizes. Interestingly, Dyce and O'Connor (1998) with a sample of 614 students provided one of the better estimates albeit using an unusually small p-value of predictor entry of .0001 in their stepwise regressions. They obtained a rather small estimate of incremental prediction with an average  $R^2$  of .35 for factors and .39 for facets.

In the context of personality and work, there has been much conceptual debate about the importance of factors and facets (e.g., Ashton, 1998; Ones & Viswesvaran, 1996; Salgado et al., 2013). Christiansen and Robie (2011) provide a clear articulation of several issues around sample sizes and types of r-squared. Some studies, mostly examining job performance, have relied on judge selected facets or highest correlating facets, some only focus on conscientiousness and its facets, and others do not report facet regression models at all. Thus, there is a need for additional empirical assessment of  $\Delta\rho^2$  using appropriate methods, all Big 5 factors, and large sample sizes.

In summary, the main aim of the current research was to evaluate and make recommendations about methods for assessing the incremental variance explained by personality facets over factors. Such methods are important in order to enable researchers to provide a fair assessment of the benefits gained by facet-level analysis in light of the substantial loss of parsimony. Specifically, this study led us to recommend the use of Olkin-Pratt adjusted r-squared along with DAB confidence intervals for estimating incremental variance explained by facets. We have also provided an R package to assist researchers to implement these methods in future studies. By reappraising the existing literature using these recommended methods, and applying them to future studies, we hope that better meta-analytic estimates can be formed about the incremental value of facet-level analysis.

### 10. References

- Alf Jr, E. F., & Graf, R. G. (1999). Asymptotic confidence limits for the difference between two squared multiple correlations: A simplified approach. *Psychological Methods*, 4(1), 70.
- Algina, J., Keselman, H., & Penfield, R. D. (2007). Confidence intervals for an effect size measure in multiple linear regression. *Educational and psychological measurement*, 67(2), 207-218.
- Anglim, J., & Grant, S. (in press). Predicting Psychological and Subjective Well-Being from Personality: Incremental Prediction from 30 Facets over the Big 5. *Journal of Happiness Studies*. doi: 10.1007/s10902-014-9583-7
- Ashton, M. C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organizational Behavior*, 19(3), 289-303. doi: 10.1002/(SICI)1099-1379(199805)19:3<289::AID-JOB841>3.0.CO;2-C
- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality*, 29(4), 432-442. doi: 10.1006/jrpe.1995.1025
- Ashton, M. C., Paunonen, S. V., & Lee, K. (2014). On the validity of narrow and broad personality traits: A response to. *Personality and Individual Differences*, 56, 24-28.
- Bagby, R. M., Costa, P. T., Widiger, T. A., Ryder, A. G., & Marshall, M. (2005). DSM - IV personality disorders and the Five - Factor Model of personality: a multi - method examination of domain - and facet - level predictions. *European Journal of Personality*, 19(4), 307-324. doi: 10.1002/per.563
- Barten, A. P. (1962). Note on unbiased estimation of the squared multiple correlation coefficient. *Statistica Neerlandica*, 16(2), 151-164.
- Baudin, N., Aluja, A., Rolland, J.-P., & Blanch, A. (2011). The role of personality in satisfaction with life and sport. *Behavioral Psychology/Psicologia Conductual*, 19(2), 333-345.
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of personality*, 80(1), 219-251.
- Christiansen, N. D., & Robie, C. (2011). Further consideration of the use of narrow trait scales. *Canadian Journal of Behavioural Science*, 43(3), 183-194. doi: 10.1037/a0023069
- Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, 64(1), 21-50. doi: 10.1207/s15327752jpa6401\_2
- de Jong, P. F. (1999). Hierarchical regression analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(2), 198-211.
- Derksen, S., & Keselman, H. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2), 265-282. doi: 10.1111/j.2044-8317.1992.tb00992.x
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71-75. doi: 10.1207/s15327752jpa4901\_13
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91(1), 40-57. doi: 10.1037/0021-9010.91.1.40
- Dyce, J. A., & O'Connor, B. P. (1998). Personality disorders and the five-factor model: A test of facet-level predictions. *Journal of Personality Disorders*, 12(1), 31-45. doi: 10.1521/pedi.1998.12.1.31

- Ekehammar, B., & Akrami, N. (2007). Personality and prejudice: From Big Five personality factors to facets. *Journal of Personality*, 75(5), 899-926. doi: 10.1111/j.1467-6494.2007.00460.x
- Ezekiel, M., & Fox, K. A. (1959). *Methods of correlation and regression analysis: Linear and curvilinear*. Oxford, England: John Wiley.
- Fox, J. (2002). *An R and S-Plus companion to applied regression*. Thousand Oaks, CA: Sage.
- Fruyt, F., Clercq, B. J., Wiele, L., & Heeringen, K. (2006). The validity of Cloninger's Psychobiological Model versus the Five - Factor Model to predict DSM - IV personality disorders in a heterogeneous psychiatric sample: Domain facet and residualized facet descriptions. *Journal of Personality*, 74(2), 479-510. doi: 10.1111/j.1467-6494.2006.00382.x
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, 7, 7-28.
- Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer.
- Hastings, S. E., & O'Neill, T. A. (2009). Predicting workplace deviance using broad versus narrow personality variables. *Personality and Individual Differences*, 47(4), 289-293. doi: 10.1016/j.paid.2009.03.015
- Judge, T. A., Klinger, R., Simon, L. S., & Yang, I. W. F. (2008). The contributions of personality to organizational behavior and psychology: Findings, criticisms, and future research directions. *Social and Personality Psychology Compass*, 2(5), 1982-2000.
- Kromrey, J. D., & Hines, C. V. (1995). Use of empirical estimates of shrinkage in multiple regression: a caution. *Educational and Psychological Measurement*, 55(6), 901-925. doi: 10.1177/0013164495055006001
- Leach, L. F., & Henson, R. K. (2007). The use and impact of adjusted R<sup>2</sup> effects in published regression research. *Multiple Linear Regression Viewpoints*, 33(1), 1-11.
- Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology*, 55(4), 675. doi: 10.1037/0022-3514.55.4.675
- O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43(5), 971-990.
- O'Neill, T., Paunonen, S., Christiansen, N., & Tett, R. (2013). Breadth in personality assessment: Implications for the understanding and prediction of work behavior. *Handbook of personality at work*, 299-332.
- Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *The Annals of Mathematical Statistics*, 201-211.
- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17(6), 609-626. doi: 10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K
- Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, 97(2), 307.
- Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology*, 74, 538-556. doi: 10.1037/0022-3514.74.2.538
- Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, 81(3), 524-539. doi: 10.1037/0022-3514.81.3.524

- Paunonen, S. V., Haddock, G., Forsterling, F., & Keinonen, M. (2003). Broad versus narrow personality measures and the prediction of behaviour across cultures. *European Journal of Personality*, 17(6), 413-433.
- Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior*, 20(3), 389-405.
- Quevedo, R. J. M., & Abella, M. C. (2011). Well-being and personality: Facet-level analyses. *Personality and Individual Differences*, 50(2), 206-211. doi: 10.1016/j.paid.2010.09.030
- 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.
- Rothmann, S., & Coetzer, E. (2002). The relationship between personality dimensions and job satisfaction. *Business Dynamics*, 11(1), 29-42.
- Salgado, J. F., Moscoso, S., & Berges, A. (2013). Conscientiousness, its facets, and the prediction of job performance ratings: Evidence against the narrow measures. *International Journal of Selection and Assessment*, 21(1), 74-84.
- Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and life satisfaction: A facet-level analysis. *Personality and Social Psychology Bulletin*, 30(8), 1062-1075. doi: 10.1177/0146167204264292
- Siegler, I. C., & Brummett, B. H. (2000). Associations among NEO personality assessments and well-being at midlife: Facet-level analyses. *Psychology and Aging*, 15(4), 710-714. doi: 10.1037/0882-7974.15.4.710
- Stephan, Y. (2009). Openness to experience and active older adults' life satisfaction: A trait and facet-level analysis. *Personality and Individual Differences*, 47(6), 637-641. doi: 10.1016/j.paid.2009.05.025
- Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior*, 24(3), 335-356. doi: 10.1002/job.191
- Yin, P., & Fan, X. (2001). Estimating R<sup>2</sup> shrinkage in multiple regression: A comparison of different analytical methods. *The Journal of Experimental Education*, 69(2), 203-224. doi: 10.1080/00220970109600656

## 11. Appendix: Additional Estimators

Three additional approaches in rough order of popularity include stepwise regression, best correlating facet selection, and cross-validated regression. First, many studies have used a variant of stepwise regression to estimate  $\Delta\rho^2$ , typically using  $R_{(\text{facets})}^{2(\text{stepwise})} - R_{(\text{factors})}^{2(\text{stepwise})}$ . In general, there is a large statistical literature that is highly critical of stepwise regression. For example, Harell (2001) catalogues many problems including that the standard errors for coefficients are inaccurate, predictor selection has a degree of arbitrariness particularly in the presence of collinearity, and in general it discourages thinking about model development. Nonetheless, while not as good as using adjusted r-squared with direct entry, there are several reasons to expect that estimates of  $\Delta\rho^2$  provided in the literature using stepwise regression may be relatively unbiased. Second, another commonly used data driven approach involves estimating  $\rho_{\text{facets}}^2$  using a set of predictors composed of the one facet from each factor that correlates most with the criterion (e.g., Hastings & O'Neill, 2009). We label this the best facets approach. This approach is not suited to

estimating  $\Delta\rho^2$ . Finally cross-validated r-squared is also not suitable as estimator of  $\Delta\rho^2$ .

### 11.1. Stepwise Regression

Asymptotically, all predictors with non-zero coefficients would be retained in a stepwise model with 100% power, and as such asymptotically the sample regression equation is the same as the population regression equation. Thus, asymptotically  $R^{2(\text{stepwise})} = \rho^2$ . However, with a finite sample size, two processes operate in opposing directions. First,  $R^{2(\text{stepwise})}$  is a positively biased estimate of  $\rho^2$  for a regression containing the same predictors selected by the stepwise procedure. However, over the population of possible studies and stepwise models, the predictors that are excluded from the stepwise regression have a non-zero probability of having non-zero population coefficients. As such the  $\rho^2$  with all predictors will tend to be greater than the  $\rho^2$  including just the stepwise predictors. So by excluding those predictors,  $R^{2(\text{stepwise})}$  would be biased lower. Thus, the direction and degree of bias in  $R^{2(\text{stepwise})}$  depends on which of the above two forces is stronger. The strength of these two forces can be modulated by manipulating the criterion p-value of the F test for entry of predictors, which we denote  $p_{\text{enter}}$ . Thus, if  $p_{\text{enter}} = 1.0$  all predictors are included and  $R^{2(\text{stepwise})}$  is positively biased. If  $p_{\text{enter}}$  is sufficiently small (e.g.,  $p_{\text{enter}} = 1.0^{-100}$ ), all predictors would be excluded, and  $R^{2(\text{stepwise})}$  is negatively biased. Given that the bias is a monotonic function of  $p_{\text{enter}}$ , there would be a value of  $p_{\text{enter}}$  for a given dataset that would lead to an unbiased estimate, and there may be a range of  $p_{\text{enter}}$  where the bias is sufficiently small to not be a problem for applied considerations.

However, it is unclear what value of  $p_{\text{enter}}$  would yield an unbiased estimate. Furthermore, the optimal  $p_{\text{enter}}$  is likely to depend on properties of the data. This confusion partially explains the wide range of  $p_{\text{enter}}$  values used in the literature including .10 (Schimmack et al., 2004), .05 (Baudin et al., 2011), .01 (Ekehammar & Akrami, 2007), and .0001 (Dyce & O'Connor, 1998). It also raises the issue of whether a different value of  $p_{\text{enter}}$  is required when there are for example five factors versus 30 facet predictors. Paunonen (1998) used a more stringent  $p_{\text{enter}}$  for facets than factors in an attempt to compensate for the perceived greater chance of including a predictor that makes no incremental prediction in the population. Given this confusion about optimally unbiased  $p_{\text{enter}}$  this raises issues with the use of stepwise regression for the purpose of examining incremental prediction of facets over factors. It is important to



have a sense of the size of the bias when evaluating the existing literature. However, we conclude as did Derksen and Keselman (1992) that adjusted r-squared is more suited to the task of providing an unbiased estimate of  $\rho^2$ .

### 11.2. Best facets

The best facets approach involves comparing the prediction by the five factors with the prediction of five facets, where the one facet that has the highest sample correlation with the criterion is included in the predictor set. It appears to have been motivated by a desire to equate the number of predictors across factor and facet regressions. Let us label the resulting estimates  $R^{2(\text{bestfacet})}$ . Asymptotically,  $R^{2(\text{bestfacet})}$  is less than or equal to  $\rho^2$  with all facet predictors. Specifically,  $R^{2(\text{bestfacet})} < \rho^2$  when any of the facets not included have non-zero coefficients in the population, which will almost always be the case.  $R_{\text{adj}}^{2(\text{bestfacet})}$  will also often be a biased high estimate of  $\rho^{2(\text{bestfacet})}$ , because the choice of predictors will vary from sample to sample with some probability and the facets that are selected will be those, which happen to provide greater correlations due to sampling error. The degree of this bias will be reduced as sample size increases and to the extent that there is a clear distinction between the largest and next largest facet correlates. Particularly small sample sizes may compensate for the fact that  $\rho^{2(\text{bestfacet})} < \rho^2$ . Nonetheless, the exact sample size where an unbiased estimate occurs is unclear, and such methods are not needed given that adjusted r-squared already provides a better estimate.

### 11.3. Cross-validated r-squared

Some researchers have proposed that cross-validated r-squared should be used to evaluate the incremental prediction of facets (e.g., Salgado et al., 2013). Finally, there are various techniques for estimating cross validated population variance explained,  $\rho_c^2$ , such as k-fold cross validation (for a review, see Kromrey & Hines, 1995) and alternative adjusted  $R^2$  formulas. Cross-validated  $R_c^2$  is not a solution to the issue of differential numbers of predictors when estimating  $\rho^2$ .  $R_c^2$  is designed to estimate  $\rho_c^2$ . It is always true that  $\rho_c^2 \leq \rho^2$ . With typical psychological sample sizes in the hundreds and values of  $R^2$  in the .1 to .7 range, expected  $\rho_c^2$  is substantially smaller than  $\rho^2$ .  $\rho_c^2$  gets closer to  $\rho^2$  as the predictions generated by the sample regression equation more closely match those from the population regression equation. On average this occurs as sample size increases.  $\rho_c^2$  is useful in applied settings where the aim is to maximize out of sample prediction. For example, this is relevant when deciding between a facet versus factor prediction model developed

in a validation sample and being applied for predicting job performance of job applicants.