R package practicalSigni Supplements Statistical Significance: Example of Racism Distorting House Values

H. D. Vinod *

March 28, 2023

Abstract

The literature on p-hacking primarily highlights its prevalence and the need to pay attention to the practical significance of each regressor. A new R package, 'practicalSigni,' reports 13 indexes for p regressors called m1 to m13. The following methods are missing from the literature. (m6) Generalized partial correlation coefficients (GPCC). (m7) extending Psychologists' "Effect Sizes" to two or more regressors. (m8, m9) Two kernel regression partial derivatives from np and NNS packages. (m10) The NNS.boost function of NNS. (m11) re-imagines the regression as a cooperative game (Shapley Value) of forecasting. Two random forest feature importance measures (m12, m13) use out-of-bin (OOB) calculations. Since regression is important for all quantitative sciences, my package offers simple few-line commands like `practicalSigni::reportRank(y,X)` to summarize the ranks by eight newer nonparametric methods (m6 to m13). An example shows why misleading p-values must

*address: H. D. Vinod, Professor of Economics, Fordham University, Bronx, New York, USA 10458. E-mail: vinod@fordham.edu. Helpful comments were made by attendees at the 75th NYSEA (New York State Economics Association) meeting on Oct. 8, 2022 and by Prof. J. Tebbs. Available at SSRN: `https://ssrn.com/abstract=4362817` JEL codes C30, C51. Keywords: Kernel regression, Standardized beta coefficients, Partial Correlation.

be supplemented by our eight methods to reveal that racism is the least significant regressor explaining Boston home values.

1 P-hacking and Practical Significance

A p-value less than 0.05 often identifies statistically significant scientific explanations of phenomena. P-hacking occurs when researchers game the estimation to achieve the 0.05 threshold and sell practically insignificant explanations as significant. The p-hacking problem is recently discussed in a special issue of the *American Statistician*, having 43 papers. Vinod (2022b) reports a p-hacking example in econometrics by Khan et al. (2020).

Let us consider a usual ordinary least squares (OLS) linear regression,

$$y_t = \alpha_0 + \Sigma_{i=1}^{p} \alpha_i x_{it} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \tag{1}$$

where $t = 1, \ldots T$. Assume that research seeks to reveal the relative importance of the p regressors in explaining y. The first measure **(m1)** of relative importance is a set of absolute values $(|\hat{\alpha}_i|, i = 0, 1, \ldots p)$. Since $(|\hat{\alpha}_i|)$ are sensitive to units of measurement, their ranking need not indicate relative importance.

The linear functional form in (1) is too restrictive and subject to specification errors. Hence we use kernel regression based on the kernel density algorithm here. The empirical cumulative distribution function (ecdf) provides $F(x)$. The density $f(x) = (dF/dx)$ is the derivative of the ecdf. One definition of a numerical derivative in calculus is the limit of the central difference. Accordingly, the density is:

$$f(x) = \lim_{h \to 0} \left(\frac{1}{h}\right) \left[F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right)\right]. \tag{2}$$

In the 1950s, Rosenblatt suggested replacing the central difference in (2) with a kernel weight function. His kernel weights are positive and must integrate (add up) to unity. A normal kernel yields convenient weights w_t $K(w_t) \sim N(0, \sigma^2)$. The standard algorithms for density estimation use $w_t = \frac{(x_t - x)}{h}$, where x is the point at which the density is evaluated, and x_t are nearby observed data points defined on an expanded grid of points. See Vinod (2021) for grid expansion details used in R.

Let X denotes a matrix of all regressors in (1). Let $f(y, X)$ denote the joint density of (y, X) variables and $f(X)$ denote the joint density of regressors, where we use a generic notation $f(.)$ for densities. The left-hand side of (1) is a conditional mean $E(y|X) = f(y, X)/f(X)$. Kernel regression applies the kernel algorithm to the two densities involved in $E(y|X)$. The resulting regression equation (in simplified notation) is

$$y_t = R(X)_t + \epsilon_t, \quad R(x) = \frac{\sum_{t=1}^{T} y_t K(w_t)}{\sum_{t=1}^{T} K(w_t)}. \tag{3}$$

Since $R(X)$ in (3) does not contain any regression parameters α_i, it is called nonparametric regression. The conditional expectation function $E(y|X) = R(X)$ based on the kernel regression algorithm traces a freehand curve very close to the data values of y. The main appeal of kernel regression is its superior fit compared to the OLS.

Our second commonly reported measure **(m2)** is the value of t-statistic which is not sensitive to units. Its absolute value ($|t_i|$) is relevant when the sign of the coefficient does not matter. For example, if the researcher wishes to show that xi increases y, and the coefficient is negative with $t_i < 0$, then the unexpected sign indicates that what the researcher wishes to establish is rejected by the data. Reporting all p-values is equivalent to t-stats, except that p-values hide "wrong" signs of coefficient estimates.

Our third measure **(m3)** overcomes the units problem of ($|\hat{\alpha}_i|$) by re-scaling all variables to have zero means and unit standard deviations. Re-scaled coefficients used to be called 'beta coefficients' or ($\hat{\beta}_i$). If the sign of the coefficient does not matter, (m3) becomes the absolute value ($|\hat{\beta}_i|$) as a measure of the importance of xi.

The commonly used fourth method **(m4)** for assessing regressor importance is the size of Pearson correlation coefficient, $|r_{y,xi}| = \sqrt{R^2}$, where R^2 is the coefficient of determination in regressing y on xi. The sign of the square root $r_{y,xi}$ is that of the covariance $cov(y, xi)$.

If the sign matters in the problem at hand, a "wrong" sign of $r_{y,xi}$ casts doubt on the researcher's claim. The first limitation of $r_{y,xi}$ is that it ignores nonlinear dependence. When one uses nonlinear nonparametric kernel regression (3), the magnitude of the R^2 depends on the regression direction (regress y on xi or regress of xi on y).

Vinod (2014) defines the square roots of the two R^2 values as elements of the asymmetric matrix R^* of generalized correlation coefficients. Kernel regression of y on xi from the square root of the R^2 denoted by $r^*_{y|xi} \neq r^*_{xi|y}$ overcomes the linearity assumption and is called **(m5)** here. The R package **generalCorr** offers a convenient function `depMeas(y,x)` for computing (m5). The second limitation of Pearson correlation ($r_{y,xi}$) is that the coefficient is between only two variables (y, xi), ignoring all others. The method (m5) overcomes linearity but continues to measure only two variables at a time.

Now we turn to the multivariate extension of ($r_{xi,xj}$). The textbook formula for partial correlation coefficient between (X_1, X_2) after removing the effect of (X_3) is:

$$r_{12;3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)}\sqrt{(1 - r_{23}^2)}}. \tag{4}$$

The numerator $(r_{12} - r_{13}r_{23})$ has the *linear* correlation coefficient between X_1 and X_2 after subtracting the *linear* effect of X_3 on both X_1 and X_2 defined in terms of regression residuals. The denominator does a normalization to obtain a scale-free correlation coefficient. Vinod (2021) generalizes a multivariate version of (4) by replacing various correlations with the square roots of R^2 from appropriate nonlinear kernel regressions.

In the context of our notation, the traditional partial correlation coefficient $r_{ij|o}$ is between two variables, xi, and xj, after removing the *linear effect* of all other variables denoted by xo. However, it is necessary to generalize $r_{ij|o}$ further by avoiding the assumed linearity leading to our method **(m6)**. Vinod (2022b) explains a generalized partial correlation coefficient (GPCC) $(y, xi|xo)$ of the indicated variables. It removes the *nonlinear effect* of all other regressors. A single-line R code for computing GPCCs is available on the web.

Section 2 details our seventh method **(m7)**, extending psychologists' effect sizes (ES). It is necessary to generalize ES to accommodate the usual regression problem with $p \geq 2$ regressors, where regressors in a kernel regression need not be dummy (0, 1) variables.

Our eighth and ninth methods of computing the practical significance use partial derivatives ($\partial y/\partial xi$) from fitted nonparametric kernel regressions using standardized data. There are many differences between algorithms for the numerical computation of these partial derivatives in two R packages. Our **(m8)** method uses the **np** package, and **(m9)** uses the **NNS** package. While np uses a chosen bandwidth for each regressor, NNS uses a dynamic bandwidth for each regressor via iterated conditional means.

Now we introduce tools from machine learning literature for overcoming the p-hacking problem. These four methods **(m10 to m13)** are computer intensive. They involve repeated iterative computations using randomization and specialized cross-validation. The underlying algorithms are explained in the literature cited by authors of various R packages used here. An intuitive explanation of all machine learning algorithms used in this paper is that they use the brute power of computers to optimize an objective function (e.g., good fit and or good out-of-sample forecast) under constraints. They all use trial and error with randomization and cross-validation.

Our tenth method measures the "importance" of a "feature" or regressor. Thus, **(m10)** refers to the results based on the R package **NNS** function `NNS.boost(.)`. It is an ensemble method using nonlinear regression based on partial moment quadrant means (NNS.reg). It stores for each random iteration a regressor list and the corresponding sum of squared prediction errors (SSPE). NNS.boost measures the importance of each regressor by the number of times (frequency) it appears among the set achieving the lowest 20% of SSPE values.

Recall that the p-hacking problem refers to the p-values of regression coefficients in (1) and (3). The method **(m11)** re-imagines the defining regression as a game of predicting y from a set of p regressors (features) along the columns of X. In the literature on cooperative game theory, Shapley Value fairly divides the payoff among p players. See Branzei et al.

(2010). We obtain a new measure of practical significance by applying a machine learning R package **ShapleyValue**. Our (m11) reports the standardized relative importance of each regressor while avoiding collinearity.

The R package **randomForest** implements the famous random forest algorithm by Breiman (2002). It also reports two measures of the "importance" of a regressor based on out-of-bin (OOB) calculations in two columns. The first column reports the percent decrease in mean squared error (MSE) upon omitting each regressor. The second column reports the mean decrease in accuracy upon omitting each regressor. The importance values in two columns are called **(m12)** and **(m13)** here.

A review of all thirteen methods suggests that each represents a unique viewpoint for the "importance rankings" of p regressors. In every real problem, the researcher needs to select a subset of relevant viewpoints from our list. The algorithms for methods (m1 to m4) are well known, but methods (m5 to m13) are not mentioned in the p-hacking literature, and some are new. More importantly, my R package **practicalSigni** allows a side-by-side comparison.

The plan for the remaining paper is as follows. We describe a new generalization of effect size in Section 2. Section 3 compares (m1 to m13) rankings side-by-side on 'mtcars' data.

2 Generalization of 'Effect Size' from Psychology

The "effect size" (ES) measurement from the Biometrics and Psychology literature can be conveniently motivated by an example in Rosenthal and Rubin (1982). They describe a medical treatment that reduced the death rate from 66% to 44%. The p-value on the death reduction exceeded 0.05, implying that the reduction was statistically insignificant. Yet the treatment saved a great many lives. See Steiger (2004), among others, for additional examples. The currently available ES measurement solves the problem with p-values subject to three limitations listed later after we establish the necessary notation. We shall see that our method (m7) generalizes the computation of ES to the general regression problem in (1).

We denote the original effect size quantifying the effect of (xi) on the outcome variable (y) as $ES_{psy}(xi)$, where the subscript (psy) indicates the psychological origin of the concept. Our extension $ES_{psyx}(xi)$ considers a set of $p \geq 2$ treatment variables (xi), $(i = 1, 2, \ldots, p)$, where the additional subscript 'x' of ES refers to extensions.

We also assume a set of control variables (xc), such as age, sex, location, etc., that influence the outcome but are outside the focus of the main research problem. In controlled experiments (xc) are placebo controls. Social scientists often employ the experimental terms 'treatments' and 'controls' though their (xi) and (xc) data are passively observed.

Following the Psychology and Biometrics literature, let us initially assume that a treatment variable (xi) is binary. That is, $xi = 1$ if the subject is treated and $xi = 0$ when the treatment is absent. If T is the sample size, let $Ti < T$ denote the subset of xi representing subjects who are treated, or those with $xi = 1$. Denote the mean treatment effect by Mxi. This is the average outcome for the subset of y items (subjects) who were treated by xi. Denote the corresponding variance by Vxi.

We assume a jointly created data generating process (DGP) having y, all xi, and control variables xc. When passively observed, DGP refers to only one set of subjects, the number of treated subjects Ti is often the same as that of control subjects Tc. Hence we generally have $Ti = Tc$, but not always. Let Mxc denote the average outcome y for the Tc control subjects. Denote the corresponding variance by Vxc.

The "effect size" in psychology is a one-sided t-statistic on the difference between two means.

$$ES_{psy}(xi) = \frac{Mxi - Mxc}{SE_i}, \tag{5}$$

where the denominator is a pooled standard error,

$$SE_i = \sqrt{(\frac{Vxi}{Ti-1} + \frac{Vxc}{Tc-1})}. \tag{6}$$

This paper extends effect size $ES_{psy}(xi)$ from psychology to more general situations having three extensions leading to our $ES_{psyx}(xi)$.

(i) $ES_{psyx}(xi)$ allows xi with $i = 1, 2, \ldots, p, (p \geq 2)$, based on regressing y on xi, and suitably defines the mean treatment effect for each xi.

(ii) $ES_{psyx}(xi)$ extends the regression to be a nonlinear kernel regression.

(iii) Our regressors xi need not be categorical variables.

Remark 1: Zero Variance Adjustment. The definition of the t-statistic of (5) or $ES_{psy}(xi)$ uses a standard textbook result on the difference between two sample means. One assumes that both random samples have sizes satisfying $(Ti \geq 30, Tc \geq 30)$, and are independent of each other. The population variances of the two samples are not assumed to be equal. The statistic (5) is not well-defined unless $SE_i \neq 0$. The central limit theorem justifies the claim that $ES_{psy}(xi)$ is an approximate scale-free Student's t statistic, unless it is degenerate.

If the xi regressor variables are converted to binary (0, 1) variables by splitting the data at the median, it can make the variances zero, implying degenerate cases. One modifies ES ratios with zero denominators by simply setting $SE_i = 1$. The degenerate ES equals $Mxi - Mxc$, or the difference between two means, subject to measurement units, not unit-free t-statistics.

Remark 2: Sign Adjustment. If the Pearson correlation, $r(y, xi) > 0$, is positive and if y refers to something desirable (e.g., profits or longevity), one wants the treatment effect means (Mxi) to exceed the (placebo) control mean (Mxc). Then, the positive effect size $ES_{psyx}(xi)$ is desirable, and negative $ES_{psyx}(xi)$ suggests the treatment is a failure and may be harmful. The opposite is true for the negative correlation $r(y, xi) < 0$. If, on the other hand, y refers to an undesirable outcome (e.g., loss or death), any successful treatment should have $ES_{psyx}(xi) < 0$. In all cases, the sign of the effect size matters. Thus the t-statistic $ES_{psyx}(xi)$ is one-sided; their ranking depends on the problem at hand. It is clear that ordering xi by effect size measured as the absolute values $|ES_{psyx}(xi)|$ is generally inappropriate. By contrast, in OLS regressions, the t-stats of regression coefficients are often two-sided, and their ordering by absolute values can be appropriate.

Now we extend $ES_{psyx}(xi)$ to be a nonlinear kernel regression using the following steps.

1. It is well known that the OLS regression framework can handle the analysis of variance methods by letting the regressors be categorical variables. If regressors are continuous, we must convert such xi regressors into binary (0, 1) dummy variables called Bxi. Our algorithm splits the data at the median.

2. Count the number Ti of Bxi values that equal unity.

3. Kernel regressing y on Bxi identifies the Ti (fitted outcome values of y) associated with $Bxi = 1$.

4. Compute their mean as Mxi and variance as Vxi for treatments.

5. Kernel regress y on all xc control regressors $(xj, j \neq i)$ and find fitted values of outcomes associated with observations having $Bxi = 1$ to make sure that their count is Tc.

6. For controls, find the mean and variance of these fitted values, denoted as Mxc and Vxc. If genuine control variable(s) are absent in the problem at hand, we assume that the control variable is a column of T ones, $xc = \iota$. Recall the intercept of the OLS regression model. Since each fitted value $\hat{y} = \bar{y}$, the mean of y, we have Mxc= $(\bar{y})$. Since the variance is degenerate, $Vxc = 0$, we must ignore the denominator in (5), making it a difference between two means but not a t-statistic.

7. Insert means and variances in (5) to compute the t-statistic $ES_{psy}(xi)$ for the difference between the two means. Of course, in degenerate cases (zero variance), we have a simple difference between two means in the numerator.

These steps are implemented in R code available on the web. When the t-stat defining $ES_{psyx}(xi)$ in (5) is degenerate due to a zero denominator, one must use only the numerator, the difference between two means $(Mxi - Mxc)$. The $ES_{psyx}(xi)$ with a degenerate denominator is no longer a t-statistic.

Effect size measurement based on relative difference while ignoring relevant magnitudes (actual temperatures) is criticized by Pogrow (2019) with a humorous example of choosing a Greenland vacation by comparing it to Antarctica. He suggests the importance of including relevant measurements, not just relative values. Some researchers suggest confidence intervals around estimated effect sizes. A bootstrap confidence interval around our $ES_{psyx}(xi)$ is easy to compute.

The following section illustrates a real-world application based on fuel economy data for 32 cars called 'mtcars' always available in R.

3 Fuel Economy Regressor Orderings for m1 to m13

This section illustrates the ordering of p=3 regressors by their "practical significance" based on thirteen viewpoints. The 'mtcars' data in R is from the 1974 US magazine called *Motor Trend*. It has 32 observations for 11 variables. We choose miles per gallon (mpg) as the dependent variable y, and the regressors are the number of cylinders (cyl), horsepower (hp), and weight (wt). We assign the rank value 1 to the most important regressor, while assigning the rank value p to the least important regressor. The index value signs are adjusted to be consistent with our ranking convention.

Table 1 reports the importance index values of three regressors based on linear and or bivariate methods (m1) to (m5). The signs have been adjusted so that a larger importance index suggests greater practical importance in explaining the dependent variable y (mpg).

Table 2 reports the corresponding rank ordering of regressors. The column for m2 reports traditional t-stat ranking where wt has the largest t-stat implying wt is the statistically most significant regressor. The average rank in the last column also suggests wt as the most important, cyl as the second most important, and hp as the least important (consistent with t-stat values)

Table 3 reports the importance index values of three regressors based on more modern nonlinear, nonparametric, and or multivariate (comprehensive) methods (m6) to (m13). As before, the signs have been adjusted so that a larger importance index suggests greater practical importance in explaining the dependent variable y (mpg). Table 4 reports the corresponding rank ordering of regressors. The average rank in the last column also suggests wt as the most important, cyl as the second most important, and hp as the least important (similar to the t-stat magnitudes).

Table 1: Importance indexes of (cyl, hp, wt) for methods (m1 = OLS), to (m5 =generalized correlation coefficient) after sign adjustment (larger the better)

	$\hat{\alpha}_i$	t-stat	$\hat{\beta}_i$	$r_{y,xi}$	depMeas(y,xi)
	m1	m2	m3	m4	m5
cyl	0.942	1.709	0.279	0.852	0.943
hp	0.018	1.519	0.205	0.776	0.938
wt	3.167	4.276	0.514	0.868	0.917

Table 2: Ranking by linear and or bivariate criteria m1 to m5 and average rank for the five methods in the last column

	m1	m2	m3	m4	m5	avrank
cyl	2	2	2	2	1	1.80
hp	3	3	3	3	2	2.80
wt	1	1	1	1	3	1.40

Table 3: Sign adjusted importance values of regressors by multivariate nonlinear methods (larger number means greater practical importance in explaining y)

	GPCC	ES_{psyx}	np	NNS	boost	Shapley	forest1	forest2
	m6	m7	m8	m9	m10	m11	m12	m13
cyl	0.0019	4.8739	5.1359	0.4291	0.3333	0.3362	16.75	285.1
hp	0.3886	4.5630	5.8848	0.2562	0.3333	0.2608	16.09	331.3
wt	0.4812	4.3038	5.7036	0.2633	0.3333	0.4030	15.43	368.9

Table 4: Ranking by nonlinear and or multivariate criteria m6 to m13 and average rank for the eight methods in the last column

	m6	m7	m8	m9	m10	m11	m12	m13	avrank
cyl	3	1	3	1	2	2	1	3	2
hp	2	2	1	3	2	3	2	2	2.10
wt	1	3	2	2	2	1	3	1	1.90

This paper argues that researchers can avoid inadvertent p-hacking if they supplement p-values by evaluating practical significance. We focus particular attention on readily measured newer methods (m6 to m13) that are more comprehensive than OLS in the sense that they allow for multivariate nonlinear nonparametric relations.

Each method (m1) to (m13) highlights different aspects, and considering several methods plus graphics will mitigate the p-hacking problem.

The p-hackers focus exclusively on the ordering of regressors by their statistical significance (or our method m2 using the t-stats). Other comprehensive viewpoints (m6 to m13) reveal alternative measures of the practical importance of a regressor worthy of consideration.

4 Determinants of Home Prices in Boston

Now we consider Hedonic Prices of Census Tracts in Boston based on a cross-section of 506 owner–occupied homes in 1970. The sample size of 506 is much larger than the 30 cars studied in Section 3. Moreover, with ten regressors, we can study their ranking providing insights for housing market choices more meaningfully than the ranking of three engineering variables.

We are interested in ranking the determinants of median values of these homes to assess the extent of racism in Boston some fifty years ago. The data are available in textbooks and an R package 'Ecdat.'

The names and descriptions of variables are

1. mv: median value of owner–occupied homes (the dependent variable)

2. crim: crime rate (not racist)

3. indus: proportion of nonretail business acres (not racist)

4. nox: annual average nitrogen oxide concentration in parts per hundred million (not racist)

5. rm: average number of rooms (not racist)

6. age: proportion of owner units built prior to 1940 (not racist)

7. dis: weighted distances to five employment centers in the Boston area (not racist)

8. rad: index of accessibility to radial highways (not racist)

9. tax: full value property tax rate (/10,000) (not racist)

10. ptratio: pupil/teacher ratio (not racist)

11. blacks: proportion of blacks in the population (racist)

Similar to (1), consider OLS linear regression having $p = 10$ regressors. Let us replace the left side by its expected value allowing us to omit the error term as:

$$\begin{aligned} E(mv) &= a_0 + a_1\, crim + a_2\, indus + a_3\, nox + a_4\, rm + a_5\, age + a_6\, dis + a_7\, rad \\ &\quad + a_8\, tax + a_9\ ptratio + a_{10}\ blacks. \end{aligned} \tag{7}$$

The adjusted R^2 of this regression is 0.7062. The coefficient estimates, with the usual details, such as t-stats and p-values, are in Table 5. Note that all regressors are statistically significant and that the p-values in the column entitled $\Pr(>|t|)$ are all less than 0.05, except for the regressor *indus* measuring the proximity of the house to industrial businesses.

Sign Adjustment: Table 5 has an additional column entitled $r_{y,xi}$ for the Pearson correlation coefficient of mv and each regressor. The sign of the correlation coefficient in the last column matters for our evaluation of various criteria for their practical significance.

The correlation between *mv* and *crim* is -0.528. We assume that the correct sign is also negative, implying that the median value *mv* of a house will decrease if crime increases. The correlation between *mv* and *rm* is $+0.642$. We assume that the correct sign is also positive, implying that the median value *mv* of a house will increase if the number of rooms in the house, *rm*, is larger. Our indexes for practical significance are adjusted for the sign so that a larger index value suggests greater practical significance.

Wrong Signs of *dis* and *rad* coefficients: Table 5 reports OLS coefficients with negative signs for *dis*, distance from job locations, and for *rad*, for accessibility to highways. According to the "correct" signs appearing in the last column for correlation coefficients, the house value increases for house locations near highways and job locations. Yet OLS coefficients are negative. For easy reference, we regard the implausible signs of OLS estimated coefficients attached to the regressors *dis* and *rad* as simply "wrong."

The largest five correlation coefficient magnitudes among our ten regressors are (0.735, 0.761, 0.778, 0.792, and 0.851). These being rather large, one suspects that collinearity is causing the "wrong" signs. Vinod (2022a) (ch.1) studies the collinearity problem from the rigorous perspective of numerical mathematicians, who define a "singular value decomposition" (R function *svd*) for any data matrix. Our data matrix X has $T = 506$ rows and $p = 10$ columns for the ten regressors, leading to the singular values $d_i, i = 1, 2, \ldots 10$. The conditioning of a matrix is defined by the condition number $K^{\#}(X) = max(d_i)/min(d_i)$. Numerical mathematicians view collinearity as a consequence of "too large" $K^{\#}(X)$ implying ill-conditioned data. Section 1.9.2.2 of Vinod (2022a) states the following *Rule of Thumb*, test, $K^{\#}(X) > 10p$, to conclude that the condition number is "too large" for regression applications. For our data, $K^{\#}(X) = 5283.958 > 10p = 100$ holds. Hence, we do have collinear data. However, using a ridge regression remedy discussed in Vinod (1978) is not an appropriate option here.

If we are interested in Boston racism in 1970, the coefficient a_{10} of *blacks* for the proportion of blacks in the population is statistically significant, having the p-value 0.0000. One may be tempted to conclude that racism played an important role in determining home prices. Our detailed analysis of this example will show that statistical significance is not the whole story, and should be supplemented by the measurement of "practical significance."

Table 5: OLS estimates for hedonic regression of median value of homes in Boston in 1970

	Estimate	Std. Error	t value	Pr(>|t|)	$r_{y,xi}$
(Intercept)	10.5390	0.1692	62.29	0.0000	
crim	-0.0137	0.0015	-9.20	0.0000	-0.528
indus	-0.0011	0.0028	-0.40	0.6901	-0.542
nox	-0.0078	0.0014	-5.74	0.0000	-0.496
rm	0.0188	0.0012	15.06	0.0000	0.642
age	-0.0031	0.0006	-5.34	0.0000	-0.453
dis	-0.2443	0.0388	-6.30	0.0000	0.406
rad	0.0937	0.0225	4.17	0.0000	-0.435
tax	-0.0005	0.0001	-3.42	0.0007	-0.561
ptratio	-0.0397	0.0058	-6.87	0.0000	-0.502
blacks	0.6160	0.1236	4.98	0.0000	0.402

Table 6 reports numerical index values for each regressor named along the rows. The tabulated values indicate the magnitudes associated with each regressor by traditional linear and or bivariate methods (m1 to m5). The tabulated values are in the same order as in (7), adjusted for signs such that a larger index value indicates greater practical significance. The values in column entitled m1 report OLS coefficient estimates of $dis = -0.2443$ and $rad = 0.0937$. Unfortunately, the estimated OLS magnitudes have "wrong" signs conflicting with the signs of corresponding $r_{y,xi}$ in the last column. The absolute values of $dis = -0.2443$ and $rad = 0.0937$ give a wrong impression regarding their practical significance. The column m1 in Table 6 along rows for dis and rad are reported as negative to downgrade their magnitudes.

Table 7 reports the ranking of index values in columns of Table 6. It ranks variables by traditional linear and or bivariate methods. The ranks of regressors dis and rad in column m1 are respectively 10 and 9 indicating the coefficient magnitudes are practically least significant in explaining house values. The values in columns entitled (t-stat, m2) and ($\hat{\beta}_i$ m3) are also negative due to wrong OLS coefficient signs.

Recall that p-hackers focus on ranking by values along the column entitled (t-stat, m2), except that some researchers may ignore the wrong sign problem with t-stat values as indexes of significance. The ranks by m1 to m3 in Table 7 for *rad* and *dis* are the lowest (9, 10) due to the OLS wrong sign problem shared by the underlying viewpoints. The ranks in column m5 refer to the new generalized correlation coefficient, which allows for nonlinearities by using kernel regressions but is bivariate. It ignores the presence of other regressors in the model.

Table 6: Traditional linear methods m1 to m5 magnitudes to assess the practical significance

	$\hat{\alpha}_i$	t-stat	$\hat{\beta}_i$	$\lvert r_{y,xi}\rvert$	depMeas(y,xi)
	m1	m2	m3	m4	m5
crim	0.014	9.196	0.289	0.528	0.698
indus	0.001	0.399	0.019	0.542	0.727
nox	0.008	5.737	0.267	0.496	0.868
rm	0.019	15.064	0.417	0.642	0.749
age	0.003	5.335	0.212	0.453	0.614
dis	-0.244	-6.298	-0.322	0.406	0.572
rad	-0.094	-4.169	-0.201	0.435	0.551
tax	0.000	3.415	0.197	0.561	0.652
ptratio	0.040	6.874	0.210	0.502	0.656
blacks	0.616	4.983	0.138	0.402	0.513

The last columns of these tables report average ranks by various methods reported along each row. The Tables are sorted according to each table's "avrank" values.

Table 7: Ranking by traditional linear bivariate methods

	$\hat{\alpha}_i$	t-stat	$\hat{\beta}_i$	$\lvert r_{y,xi}\rvert$	depMeas(y,xi)	
	m1	m2	m3	m4	m5	avrank
rm	3	1	1	1	2	1.6
crim	4	2	2	4	4	3.2
nox	5	4	3	6	1	3.8
ptratio	2	3	5	5	5	4.0
indus	7	8	8	3	3	5.8
age	6	5	4	7	7	5.8
tax	8	7	6	2	6	5.8
blacks	1	6	7	10	10	6.8
rad	9	9	9	8	9	8.8
dis	10	10	10	9	8	9.4

Table 8 reports index values of regressor variables by newer nonlinear multivariate methods (m6 to m13). Again, tabulated regressor values follow the same order as in (7), adjusted for signs such that a larger index value indicates greater practical significance according to the viewpoint in the column title. For example, according to the generalized partial correlation coefficient (GPCC) criterion in m6, the top two regressors are neighborhood crime and pollution (*nox*). If we consider psychologists' "effect size" criterion (m7), the two top regressors are $ptratio, indus$.

Table 8: Index values by newer nonlinear multivariate methods

	GPCC	ES_{psyx}	np	NNS	boost	Shapley	forest1	forest2
	m6	m7	m8	m9	m10	m11	m12	m13
crim	0.619	0.142	6.033	0.093	0.136	0.127	25.336	12.806
indus	-0.207	0.194	-0.720	0.125	0.045	0.075	12.741	5.127
nox	0.592	0.190	2.630	0.113	0.091	0.076	25.122	13.131
rm	0.340	0.184	0.235	-0.071	0.045	0.304	53.674	21.384
age	0.183	0.177	0.283	0.135	0.136	0.061	21.563	3.885
dis	0.036	0.145	0.142	-0.110	0.091	0.045	18.899	7.087
rad	-0.265	0.162	-0.030	0.126	0.136	0.045	9.463	1.236
tax	0.038	0.179	0.606	0.148	0.136	0.089	11.470	4.474
ptratio	0.407	0.210	-0.049	0.109	0.045	0.117	17.351	8.366
blacks	0.229	0.019	-0.275	-0.107	0.136	0.059	15.712	3.567

Table 9 reports the ranking of variables by newer nonlinear multivariate methods based on the index values in Table 8, where the regressors are rearranged according to the average ranks in the last column. We find that newer multivariate nonlinear criteria (m6 to m13) regard $crim, nox$ and rm as the three most important regressors. The size of the home measured by rm is less important than neighborhood crime and pollution.

Table 9: Ranking by newer nonlinear multivariate methods

	GPCC	ES_{psyx}	np	NNS	boost	Shapley	forest1	forest2	
	m6	m7	m8	m9	m10	m11	m12	m13	avrank613
crim	1.0	9.0	1.0	7.0	3.0	2.0	2.0	3.0	3.5
nox	2.0	3.0	2.0	5.0	6.5	5.0	3.0	2.0	3.6
rm	4.0	4.0	5.0	8.0	9.0	1.0	1.0	1.0	4.1
tax	7.0	5.0	3.0	1.0	3.0	4.0	9.0	7.0	4.9
age	6.0	6.0	4.0	2.0	3.0	7.0	4.0	8.0	5.0
ptratio	3.0	1.0	8.0	6.0	9.0	3.0	6.0	4.0	5.0
indus	9.0	2.0	10.0	4.0	9.0	6.0	8.0	6.0	6.8
dis	8.0	8.0	6.0	10.0	6.5	9.0	5.0	5.0	7.2
rad	10.0	7.0	7.0	3.0	3.0	10.0	10.0	10.0	7.5
blacks	5.0	10.0	9.0	9.0	3.0	8.0	7.0	9.0	7.5

Let us compare the importance ranking of our ten regressors in Table 7 based on five older (m1 to m5, linear bivariate) criteria with Table 9 summarizing newer nonlinear and multivariate (m6 to m13) viewpoints. The most important regressor is the size of the home (rm) by older criteria and crime by newer criteria. The racist variable, *blacks*, is ranked third from the bottom in Table 7 for older methods. By contrast, all nine regressors are considered more important in determining the home values than the racist variable *blacks* in Table 9 for newer methods m6 to m13.

A naive comparison of p-values in Table 5 suggests all regressors have nine comparably "small" p-values 0.0000 (rounded to four digits), and could not have afforded the following insight about the role of racism. We demonstrate that the housing prices in Boston in 1970 were determined *more* by regular market forces, not racist prejudice.

4.1 R code for the housing example

The following code implementation on my home computer took all night to implement. The options m6 and m10 are time-consuming. There is an option to not to bother with computing m6 and m10 to save computer time by setting yes13[6]=0 and yes13[10]=0.

```
library(practicalSigni)
options(np.messages=FALSE)
library(Ecdat)
attach(Hedonic)
bigx=cbind(crim,indus,nox,rm,age,dis,rad,tax,ptratio,blacks)
r1=reportRank(y=mv,bigx,verbo=TRUE) #this can take 10 hours

#following unneeded code merely sorts the results.
vmtx15=r1$v15
vmtx613=r1$v613
rmtx15=r1$r15
rmtx613=r1$r613
```

```
so15= sort_matrix(rmtx15,ncol(rmtx15))

so613= sort_matrix(rmtx613,ncol(rmtx613))
```

5 Final Remarks

Recent literature focuses on the need to avoid the p-hacking problem, Wasserstein et al. (2019), arising from excess reliance on regression coefficient p-values (t-stat) while ignoring "practical significance." Since linearity and normality are strong assumptions, we argue for nonlinear nonparametric kernel regressions to supplement the OLS common in most p-hacking examples. While many papers demonstrate the existence of the p-hacking problem, very few provide algorithms (software) to solve the problem. Vinod (2022b) proposes two methods (m6) using generalized partial correlation coefficients (GPCC) and a partial derivative (m8) using the 'np' package to solve the p-hacking problem.

Some Psychometrics and Biometrics researchers have long ago proposed an intuitive method of measuring practical significance. They consider "effect sizes" $ES_{psy}(xi)$ instead of t-stats. This paper describes a new method (m7 and software algorithm steps) for what we denote as $ES_{psyx}(xi)$. The additional subscript 'x' suggests an extension by admitting $p \geq 2$, possibly continuous regressors, and nonlinear kernel regressions.

Among new methods in this paper, (m9) is analogous to (m8) using the 'NNS' package. This paper provides algorithms for implementing four new measures of practical significance (m10 to m13) using machine learning algorithms. The machine learning methods can be sensitive to random seeds. The ultimate choice must depend on the scientific context.

This paper reports a real-world example of fuel economy regression. The average rank computed over eight methods is a good summary of various viewpoints deserving serious consideration. We also compute the average rank (avrank) over the eight newer methods (m6 to m13). Our second example using data on determinants of home prices shows that the home values in Boston in 1970 were determined by normal market forces and not racist prejudice. A naive comparison of OLS p-values in Table 5 displays that eight of ten slope p-values are very small (=0.0000). Since the racism regressor *blacks* is one of eight regressors sharing the small p-value, racism appears to be highly statistically significant. By contrast, further analysis of practical significance shows *blacks* as the least important regressor implying that racism was less critical. This example demonstrates that practitioners should consider supplementing their p-values with comprehensive tools readily computed by simple R commands (section 4.1).

References

Branzei, R., Branzei, O., Gök, S., and Tijs, S. (2010), "Cooperative interval games: a survey," *Central European Journal of Operations Research*, 387, 397–411.

Breiman, L. (2002), *Manual On Setting Up, Using, And Understanding Random Forests V3.1*, URL `https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf`.

Khan, A., Khan, M. I., and Bhutto, N. A. (2020), "Reassessing the Impact of Risk Management Capabilities on Firm Value: A Stakeholders Perspective," *Sukkur IBA Journal of Management and Business*, 6(2), 81–98.

Pogrow, S. (2019), "How Effect Size (Practical Significance) Misleads Clinical Practice: The Case for Switching to Practical Benefit to Assess Applied Research Findings," *The American Statistician*, 73, 223–234.

Rosenthal, R. and Rubin, D. B. (1982), "A simple, general purpose display of magnitude of experimental effect," *Journal of educational psychology*, 74(2), 166–169.

Steiger, J. H. (2004), "Beyond the F Test: Effect Size Confidence Intervals and Tests of Close Fit in the Analysis of Variance and Contrast Analysis," *Psychological Methods*, 9(2), 164–182.

Vinod, H. D. (1978), "A survey of ridge regression and related techniques for improvements over ordinary least squares," *Review of Economics and Statistics*, 60, 121–131.

— (2014), "Matrix Algebra Topics in Statistics and Economics Using R," in "Handbook of Statistics: Computational Statistics with R," , eds. Rao, M. B. and Rao, C. R., New York: North-Holland, Elsevier Science, vol. 34, chap. 4, pp. 143–176.

— (2021), "Generalized, Partial and Canonical Correlation Coefficients," *Computational Economics*, 59(1), 1–28.

— (2022a), *Hands-on Intermediate Econometrics Using R: Templates for Learning Quantitative Methods and R software, Second edition*, Hackensack, NJ: World Scientific, ISBN 978-981-125-617-2, https://doi.org/10.1142/12831, URL `https://www.worldscientific.com/worldscibooks/10.1142/12831`.

— (2022b), "Kernel regression coefficients for practical significance," *Journal of Risk and Financial Management*, 15, 1–13, URL `https://doi.org/10.3390/jrfm15010032`.

Wasserstein, R. L., Schirm, A. L., and Lazar, N. A. (2019), "Moving to a World Beyond p less than 0.05," *The American Statistician*, 73, 1–19.