Psychological Test and Assessment Modeling, Volume 58, 2016 (4), 567-592

The detection of heteroscedasticity in regression models for psychological data

Andreas G. Klein[1], *Carla Gerhard*[2], *Rebecca D. Büchner*[2], *Stefan Diestel*[3] & *Karin Schermelleh-Engel*[2]

Abstract

One assumption of multiple regression analysis is homoscedasticity of errors. Heteroscedasticity, as often found in psychological or behavioral data, may result from misspecification due to overlooked nonlinear predictor terms or to unobserved predictors not included in the model. Although methods exist to test for heteroscedasticity, they require a parametric model for specifying the structure of heteroscedasticity. The aim of this article is to propose a simple measure of heteroscedasticity, which does not need a parametric model and is able to detect omitted nonlinear terms. This measure utilizes the dispersion of the squared regression residuals. Simulation studies show that the measure performs satisfactorily with regard to Type I error rates and power when sample size and effect size are large enough. It outperforms the Breusch-Pagan test when a nonlinear term is omitted in the analysis model. We also demonstrate the performance of the measure using a data set from industrial psychology.

Keywords: Heteroscedasticity, Monte Carlo study, regression, interaction effect, quadratic effect

---

[1]*Correspondence concerning this article should be addressed to:* Prof. Dr. Andreas G. Klein, Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60629 Frankfurt; email: klein@psych.uni-frankfurt.de

[2]Goethe University Frankfurt

[3]International School of Management Dortmund Leipniz-Research Centre for Working Environment and Human Factors

Introduction

One of the standard assumptions underlying a linear model is that the errors are independently identically distributed (i.i.d.). In particular, when the errors are i.i.d., they are homoscedastic. If the errors are not i.i.d. and assumed to have distributions with different variances, the errors are said to be heteroscedastic. A linear heteroscedastic model is defined by:

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_m x_{mi} + \varepsilon_i, \;\; i = 1, \dots, n, \tag{1}$$

where the ε_i are realizations (sampled values) of error variables ε that follow a mixture distribution with normal mixing components:

$$\varepsilon \sim SZ. \tag{2}$$

For the random variables *S, Z* we assume that

$$S > 0, \quad Z \sim \mathcal{N}(0,1), \quad S \perp Z. \tag{3}$$

We are making the following regularity assumptions for S^2, the random variable that models the variances of the errors:

$$0 < E(S^2) < \infty, \quad 0 \leq var(S^2) < \infty. \tag{4}$$

Heteroscedasticity is given when $var(S^2)$ takes a positive value. In contrast, homoscedasticity holds if $var(S^2) = 0$. This definition of heteroscedasticity covers both models with a discrete and with a continuous distribution of the variances of the errors. For the random variable S^2 two types of heteroscedasticity can generally be distinguished: First, there could be a specific, parametric form of heteroscedasticity where S^2 is a function of the given predictors, such as $S^2 = exp(a_0 + a_1 x_1 + ... + a_m x_m)$. Models with this type of parametric heteroscedasticity have been investigated in the past (cf. White, 1980). Second, there could be an unspecific form of heteroscedasticity, where S^2 is entirely unrelated to the observed explanatory variables. When a linear model for a specific set of predictors is selected, heteroscedasticity of the errors may be due to different causes. For instance, in social sciences and especially in psychological research one often deals with learning mechanisms among individuals during the process of data collection. These learning mechanisms are one possible source of heteroscedastic errors, because prediction may be more accurate for subjects whose predictor scores are observed at a later stage in their development. Furthermore, reasons for heteroscedasticity could be omitted variables, outliers in the data, or an incorrectly specified model equation, for example omitted product terms. In psychological contexts product terms in regression are often related to overlooked or yet unidentified moderator variables.

In the following, we will focus on the relationship between heteroscedastic errors, model misspecification and the distribution of the regression residuals when a (possibly misspecified) regression model is fit to the data. In particular, we are interested in misspecifications resulting from omitted nonlinear terms. In psychological research, for instance, the issue of omitted interaction terms is often of particular interest, and some methodological research has been concerned with the development of efficient estimation methods in the past (cf. Dijkstra & Schermelleh-Engel, 2014; Klein & Moosbrugger, 2000; Klein & Muthén, 2007). Figure 1 gives an illustration of residuals resulting from an omitted interaction term. In the upper panel of Figure 1, an interaction model has been analyzed with a correctly specified model. In the lower panel of Figure 1, the interaction model has been analyzed with a (misspecified) linear model, where the interaction term was omitted. As can be seen, the dispersion of the residuals of the misspecified model is no longer homoscedastic.

If the model in Equation (1) is correct and if heteroscedasticity of ε holds, the regular OLS estimator for the regression coefficients still yields unbiased and consistent parameter estimates. However, the estimator is not efficient anymore, because the standard errors are biased and therefore the usual inference methods are no longer accurate (cf. Greene, 2012; White, 1980). For heteroscedastic models with correctly specified regression equation and correctly specified parametric heteroscedasticity, robust estimation methods have been developed to deal with heteroscedastic errors. Two alternatives exist that use either a weighted least squares estimator or a heteroscedasticity-consistent covariance matrix (MacKinnon & White, 1985; White, 1980). In regression analysis the distribution of the residuals depends on the heteroscedasticity of the errors and the selection of predictors to model the data. Visible heteroscedasticity may therefore often be a result of a misspecified regression model. In this case a modification of the model structure might sometimes be more useful than using a robust estimation method.

There are different ways to test for heteroscedasticity in linear regression models. One group of tests can be classified as 'model-based heteroscedasticity tests' (cf. Greene, 2003). These tests are using a specific parametric model to specify the heteroscedasticity. If this specification is incorrect, the tests may fail to identify heteroscedasticity. Three well-known statistical tests exist that are used for such parametric models: the Wald test, the likelihood ratio test, and the Lagrange multiplier test (cf. Engle, 1984; Greene, 2003; Wald, 1943). Another group of tests, which is able to detect heteroscedasticity in a more general form, can be called 'residual-based heteroscedasticity tests' (cf. Greene, 2003). Most of these tests are only available for categorical predictors (cf. Rosopa, Schaffer, & Schroeder, 2013) and are not suitable for our purposes. For both categorical and continuous predictors, two tests remain; the Breusch-Pagan and the White test. These tests can be represented by an auxiliary regression equation that uses some function of the estimated residuals as dependent variable and various functions of the proposed explanatory variables as predictor variables[1]. The well-known Breusch-Pagan test was proposed by Breusch and Pagan (1979) and by Cook and Weisberg (1983). It has been developed independently in the econometrics and statistics literature (cf. Rosopa et al., 2013). The Breusch-Pagan test tests the null hypothesis that the residuals' variances are unrelated to a set of explanatory variables versus the alternative hypothesis that the residuals' variances are a parametric function of the predictor variables. The test can be represented in an auxiliary regression form, in which the squared residuals of the proposed model are regressed on the predictors believed to be the cause of the heteroscedasticity. The common White test has been proposed by White (1980), where the squared OLS-residuals are regressed on all distinct predictors, cross products, squares of predictors, and the intercept. The test statistic is given by the coefficient of determination of the auxiliary regression multiplied by the sample size (nR^2). The statistic of the White test is chi-square distributed with degrees of freedom equal to the number of predictors in the auxiliary regression. However, these common heteroscedasticity tests do not solve the problem of detecting heteroscedasticity that is caused by omitted predictors. Therefore, Klein and Schermelleh-Engel (2010) proposed the $\mathbf{Z}_{het}$ statistic in the context of structural equation modeling. This statistic is potentially suitable to detect heteroscedasticity caused by omitted predictors in structural equation models. However, in some preliminary studies $\mathbf{Z}_{het}$ showed an undesirably low power in the detection of heteroscedasticity.

The objective of the current paper is to fill that gap of detecting unsystematic heteroscedasticity that relates to omitted predictors or yet unanalyzed nonlinear relationships and explore a basic measure of heteroscedasticity. It does not require a specification of the heteroscedasticity. Instead, the measure makes direct use of the dispersion of the squared residuals and an additional auxiliary regression is not necessary. In addition, a second goal for this paper is to find an approach that can be applied easily.

[1]Please note that tests for heteroscedasticity presented in original literature with asymptotic chi-square distributions, such as likelihood ratio, Wald or Lagrange multiplier test, are asymptotically equivalent to the auxiliary regression approach (cf. Engle, 1984).

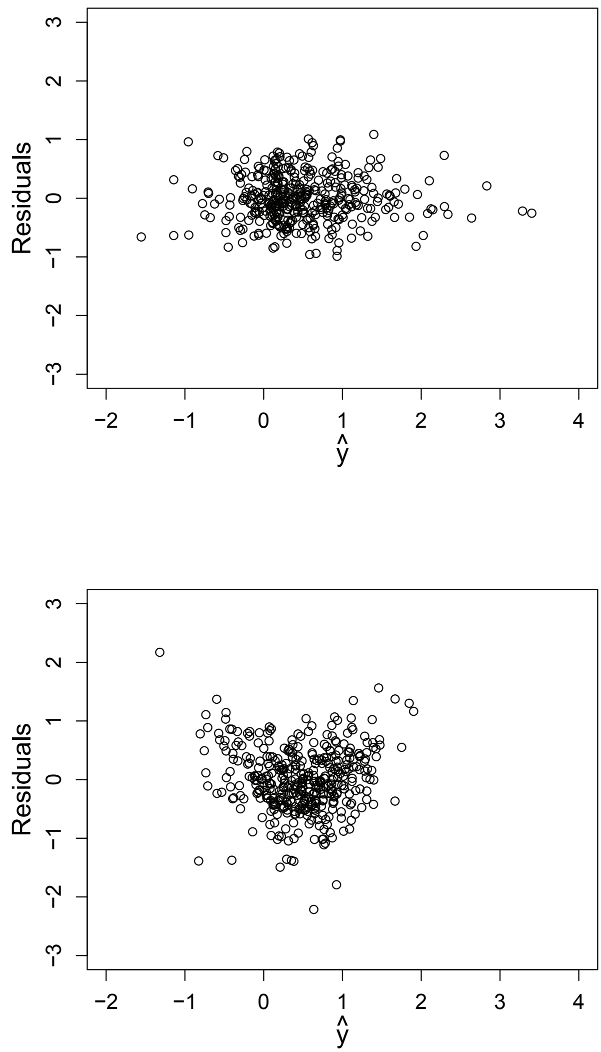

Figure 1: Scatter plot of the residuals with modeled interaction term (top) and with an omitted interaction term (bottom). Data generated for population model $y = 0.5 + 0.5x_1 + 0.3x_2 + 0.4x_1x_2 + e$, with $n = 400$ and $e \sim \mathcal{N}(0, 0.16)$.

Heteroscedasticity measure

In this section, we introduce the measure h_{het} to test for heteroscedasticity of the errors. The measure h_{het} is intended to measure a possible deviation from homoscedasticity. If the errors are heteroscedastic, they have distributions with different standard deviations, and one may then expect that the variance of the squared regression residuals e tends to be greater than it does when the residuals are homoscedastic. After conducting an OLS regression, the OLS residuals e_i $(i = 1, ..., n)$ are available for all n cases, and we have $\bar{e} = 0$. We consider

$$\begin{aligned}\frac{var(e^2)}{var(e)^2} &\approx \frac{n^{-1}\Sigma(e_i^2 - \overline{e^2})^2}{(n^{-1}\Sigma(e_i - \overline{e})^2)^2} \\ &= \frac{n^{-1}\Sigma e_i^4 - (n^{-1}\Sigma e_i^2)^2}{(n^{-1}\Sigma e_i^2)^2} \\ &= \frac{n^{-1}\Sigma e_i^4}{(n^{-1}\Sigma e_i^2)^2} - 1 \\ &= \hat{\gamma} - 1,\end{aligned} \tag{5}$$

where $\hat{\gamma}$ is the common sample-based fourth standardized moment of the residuals. $\hat{\gamma}$ is an estimator for the kurtosis of e (cf. Davidson & MacKinnon, 1993). Originally, the estimator $\hat{\gamma}$ has been formulated for independent e_i values, where it is shown that $\hat{\gamma}$ is asymptotically normally distributed (cf. Davidson & MacKinnon, 1993). Here, because the OLS residuals meet the constraint $\overline{e} = 0$, they are not independently distributed. However, we confirmed in simulation studies the asymptotic behavior of $\hat{\gamma}$ when calculated from OLS residuals for sufficiently large sample sizes ($n \geq 100$). Therefore, we can adopt the known asymptotic distribution

$$\hat{\gamma} \sim \mathcal{N}(3, 24/n) \tag{6}$$

(cf. Davidson & MacKinnon, 1993) for $\hat{\gamma}$ based on the OLS residuals. We define the measure h_{het} as

$$h_{het} := \sqrt{\frac{n}{24}}(\hat{\gamma} - 3), \tag{7}$$

so that

$$h_{het} \sim \mathcal{N}(0, 1). \tag{8}$$

In case of heteroscedastic residuals, it can be shown that, asymptotically, h_{het} takes a value greater than zero. To see this, it is sufficient to show that $\lim\limits_{n\to\infty} \hat{\gamma}$ is greater than three:

$$\begin{aligned}\lim_{n\to\infty}\hat{\gamma} &= \lim_{n\to\infty}\frac{n^{-1}\Sigma e_i^4}{(n^{-1}\Sigma e_i^2)^2} \\ &= \frac{E(\varepsilon^4)}{(E(\varepsilon^2))^2} \\ &= \frac{E(S^4)E(Z^4)}{(E(S^2)E(Z^2))^2} \\ &= 3\frac{E(S^4)}{(E(S^2))^2} \\ &= 3\frac{var(S^2)+(E(S^2))^2}{(E(S^2))^2} \\ &= 3\left[1+\frac{var(S^2)}{(E(S^2))^2}\right] > 3.\end{aligned} \tag{9}$$

Based on this result, it is indicated to use a one-tailed test for h_{het}.

The test we propose here does not make a specific assumption about what caused a possible heteroscedasticity, and it does not need a specific parametric model of the structure of heteroscedasticity. In contrast to residual-based heteroscedasticity tests, h_{het} is able to detect heteroscedasticity of the residuals that could be due to unobserved nonlinear predictor terms.

Simulation study

A Monte Carlo study was conducted with the aim of investigating the sensitivity of the measure h_{het} to respond to heteroscedasticity relating to omitted nonlinear terms. The study investigates the influence of nonlinear effect size and sample size on the performance of h_{het}. Various linear and nonlinear population models were selected for data generation and the residuals were afterwards analyzed with h_{het}. For reasons of comparability, the residuals of a model with an omitted unobserved quadratic predictor were additionally analyzed with the Breusch-Pagan test. In the following, we first introduce the population and analysis models as well as the particular design of the study; second, we present the results about the performance of h_{het}.

Population models for heteroscedasticity related to observed predictors

Different population models were used for data generation. Four population models were chosen for estimating the sensitivity of h_{het} to respond to omitted nonlinear terms. The first population model M_{LQI} was a full nonlinear model with two linear (L), two quadratic (Q) and one interaction term (I):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + e, \tag{10}$$

where $\beta_0 = .50, \beta_1 = .50$, and $\beta_2 = .30$ were held constant across all simulation conditions. For the variables x_1, x_2, and e normally distributed data were generated. The correlation between x_1 and x_2 was fixed to $r_{12} = .20$ in all simulation conditions. The variances of x_1 and x_2 were set to 1.00; the variance of e was fixed to .40 in all conditions. M_{LQI} included three nonlinear terms, the effects of these terms were varied in size correspondingly. For the first condition the effect sizes were set to $\beta_3 = \beta_4 = .10$, and $\beta_5 = .15$; for the second condition to $\beta_3 = \beta_4 = .15$, and $\beta_5 = .20$. Combined, the nonlinear terms explained between 10 % and 19 % of the variance in y.

The second population model M_{LQ} was a nonlinear model with two linear (L) and one quadratic effect (Q). M_{LQ} is the same as M_{LQI}, except for seting $\beta_4 = \beta_5 = 0$. The size of β_3 was set to .20 and .30 in two effect size conditions. The quadratic effect explains between 9 % and 19 % of the variance in y.

The third population model M_{LI} was a nonlinear model with two linear (L) and one interaction effect (I). M_{LI} is the same as M_{LQI}, except for setting $\beta_3 = \beta_4 = 0$. The size of β_5 was set to .30 and .40 in two effect size conditions, this equals an explained variance of 10 % to 18 % in y. In addition, to show the practical use of h_{het} for greater regression coefficients, M_{LI} was generated with another set of parameters. In the additional condition the regression coefficients were set to $\beta_0 = 2$, $\beta_1 = 2$, $\beta_2 = 1.2$, and to $\beta_5 = 1.2$ with the same variances and covariances as before for the error term e and the predictors x_1 and x_2. The nonlinear term explained 18 % of the variance in y.

The fourth population model M_L was a linear model with two linear (L) effects. M_L is the same as M_{LQI}, but it included no nonlinear terms after setting $\beta_3 = \beta_4 = \beta_5 = 0$.

Population models for heteroscedasticity unrelated to the observed predictors

In addition to investigating the performance of h_{het} to detect omitted nonlinear terms which are related to the observed predictors, we examined the sensitivity of h_{het} to respond to heteroscedastic residuals due to unobserved predictors. For the investigation of the sensitivity of h_{het} to detect nonlinear terms of unobserved predictors, three population models were chosen: First, a quadratic model M_{LQ} with two linear (L) and one quadratic effect (Q) was used as nonlinear population model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_2^2 + e, \tag{11}$$

where $\beta_0 = .50, \beta_1 = .50$, and $\beta_2 = .30$. The variables x_1, x_2, and e were normally distributed; the correlation between x_1 and x_2 was set to $r_{12} = .20$. The size of β_4 was set to .15 and .25 in two effect size conditions.

Second, the population model M_{LI} with two linear (L) and one interaction effect (I) was used:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_1 x_2 + e. \tag{12}$$

M_{LI} is the same as M_{LQ}, except for setting $\beta_3 = 0$, and $\beta_5 = .25$ or $\beta_5 = .35$ in two effect size conditions.

Third, a linear population model M_{sL} containing only a single linear predictor (sL) was chosen:

$$y = \beta_0 + \beta_1 x_1 + e. \tag{13}$$

M_{sL} is similar to M_{LQ}, resulting from setting $\beta_2 = \beta_4 = 0$, such that M_{sL} included only a single linear predictor and no nonlinear terms.

Design

The data for the population models were generated with the R software and analyzed with the OLS estimator in R version 3.2.2 (R Core Team, 2015). For each condition R = 10,000 data sets were generated. Across all conditions, except the additional condition for M_{LI}, the sample size n was 100, 200, 400, 800 or 1,200. For the additional condition $n = 400$ was selected. For heteroscedasticity related to the observed predictors, four population models (M_{LQI}, M_{LQ}, M_{LI}, and M_L), two effects size conditions, and five sample size conditions were implemented, and each population model was analyzed as a correctly specified and as a misspecified model. For estimating h_{het} for models that contain heteroscedasticity related to unobserved predictors, three population models (M_{LQ}, M_{LI}, and M_{sL}), two effect size conditions, and five sample size conditions were implemented. For a power analysis the proportion of data sets was examined where h_{het} had values greater than the critical value at the 5 % level of a one-sized test ($z = 1.65$). Linear population models were analyzed by correctly specified models and by various overparameterized nonlinear models to study the Type I error rate. The data generated for nonlinear population models were analyzed by misspecified linear models for power analysis and by correctly specified models for Type I error analysis. As the ordinary residuals are scale dependent, some researchers recommend the use of internally or externally studentized residuals (Cook & Weisberg, 1982; Stevens, 1984). For the simulation conditions presented here, the results for ordinary residuals and for internally studentized residuals were very similar but are not reported in this article.

In the following we will provide the results for the measure h_{het}. Additionally, some results for the Breusch-Pagan test and the AIC are compared with h_{het}. The formula for the auxiliary regression in the Breusch-Pagan test is

$$\frac{e^2}{\hat{\sigma}^2} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \varepsilon, \tag{14}$$

where $\hat{\sigma}^2 = \frac{\sum e_i^2}{n}$ and ε is normally distributed with zero mean.

Results

In this section, we present the results of the simulation study. In addition to mean h_{het}-values, we report the Type I error rates and the power of h_{het} to detect omitted nonlinear terms that resulted in heteroscedasticity. The power exceeded 80 % under several conditions. The 95 % confidence interval (CI) for the error rate of a test with 5 % nominal Type I error, for a sample of 10,000 cases, is calculated as [4.57,5.43]. For h_{het}, the error rate turned out to be slightly inflated, because under some conditions the error rate was lying slightly above this range.

Heteroscedasticity related to observed predictors

The following results refer to the investigation of the influence of varying nonlinear effect size and sample size on the detection of heteroscedasticity with h_{het}.

Linear population model. For the linear population model M_L the mean h_{het}-values and Type I error rates for the different linear and nonlinear analysis models M_L, M_{LQ}, M_{LI}, and M_{LQI} are listed in Table 1. The results indicate appropriate Type I error rates close to the nominal 5 % level. Only one value was too small, and two values were slightly too high. On average the h_{het}-values tended to be slightly negative.

The probability density functions presented in Figure 2 illustrate the convergence of the distribution of h_{het} towards the standard normal distribution. For the plot, Epanechnikov kernel functions were produced. The Epanechnikov kernel was used, because it displays deviations from normality more clearly than the Gaussian kernel. The functions are shown for the linear population model M_L correctly analyzed as M_L for sample sizes

Table 1: Mean h_{het}-Values and Type I Error Rates (in Percent) as a Function of Sample Size (n) for the Linear Population Model M_L.

population model:	M_L $y=\beta_0+\beta_1x_1+\beta_2x_2$							
analysis model:	M_L $y=\beta_0+\beta_1x_1+\beta_2x_2$		M_{LQ} $y=\beta_0+\beta_1x_1+\beta_2x_2$ $+\beta_3x_1^2$		M_{LI} $y=\beta_0+\beta_1x_1+\beta_2x_2$ $+\beta_5x_1x_2$		M_{LQI} $y=\beta_0+\beta_1x_1+\beta_2x_2$ $+\beta_3x_1^2+\beta_4x_2^2+$ $\beta_5x_1x_2$	
n	Mean	Type I error	Mean	Type I error	Mean	Type I error	Mean	Type I error
100	-0.15	3.84	-0.11	4.75	-0.10	4.82	-0.09	4.97
200	-0.09	4.60	-0.08	5.33	-0.09	5.25	-0.08	5.11
400	-0.06	5.00	-0.05	5.62	-0.05	5.57	-0.06	5.10
800	-0.06	5.16	-0.04	5.07	-0.03	5.23	-0.05	5.52
1,200	-0.04	5.20	-0.05	5.18	-0.04	5.20	-0.04	5.39

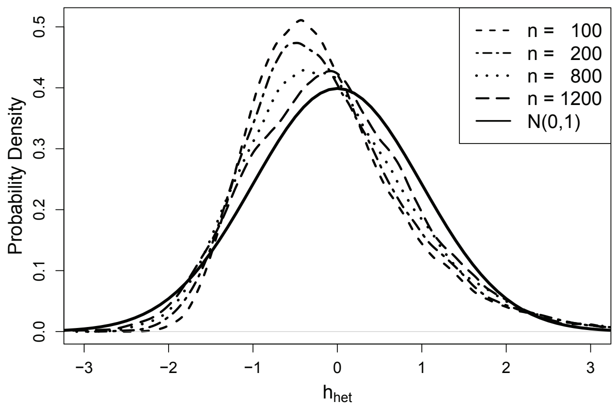

Figure 2: Estimated probability density function of h_{het} for the linear Population Model M_L correctly analyzed as M_L. The density functions were estimated using an Epanechnikov kernel function.

$100, 200, 800$ and $1,200$. The density function for $n = 1,200$ is close to the $\mathcal{N}(0,1)$-density and has a kurtosis of 0.36 and a skewness of 0.42. The kurtosis for $n = 100$ is 3.20, whereas the skewness is 1.22. Both kurtosis and skewness decrease with greater sample size. In the critical part of the distribution, the right hand tail, there were only small deviations from the standard normal curve. We note that for small samples the density curve for $h_{het} > 1.645$ is slightly above or below the ideal normal density. As a consequence, the Type I error does not deviate much from .05 (see Table 1).

Nonlinear population model. In Table 2 the results for the quadratic population model M_{LQ} are presented. In addition, the power of h_{het} is listed, where h_{het} correctly indicates the presence of heteroscedasticity in the residuals of the linear analysis model M_L. It appears that the Type I error rates were close to the nominal 5 % level for all sample sizes, where one value was too high. A desirable power of 80 % was exceeded at sample size $n = 600$ when the nonlinear effect size was $\beta_3 = .30$. For a quadratic effect size of $\beta_3 = .20$ a power of 80 % was not reached for the listed values. Additional simulations indicate a required sample size of $n \geq 2,700$ (not listed in Table 2).

The results for the population model M_{LI} can be seen from Table 3. The Type I error rates were again close to their nominal 5 % levels in all conditions, four values were just outside the 95 % CI bounds. A power close to 80 % was reached for sample size of 1,200 and interaction effect size of $\beta_5 = .40$. For a small effect size ($\beta_5 = .30$) a sample size

Table 2: Mean h_{het}-Values, Type I Error Rates (in Percent), and Power (in Percent) as a Function of Sample Size (n) and Quadratic Effect Size for the Quadratic Population Model M_{LQ}.

population model:	M_{LQ} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2$							
analysis model:	M_{LQ} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2$				M_L $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$			
nonlinear pop. parameter:	$\beta_3 = .20$		$\beta_3 = .30$		$\beta_3 = .20$		$\beta_3 = .30$	
n	Mean	Type I error	Mean	Type I error	Mean	Power	Mean	Power
100	-0.12	4.59	-0.12	4.76	0.20	10.02	1.07	26.17
200	-0.10	4.87	-0.08	5.31	0.62	18.47	2.26	45.75
400	-0.07	5.09	-0.06	5.40	1.07	27.35	3.87	70.77
800	-0.05	5.29	-0.04	5.09	1.69	41.52	5.98	91.07
1,200	-0.03	5.47	-0.05	5.43	2.19	54.23	7.53	97.41

of at least $n = 4,000$ is needed (not listed in Table 3). For the population model M_{LI}, the *AIC* was in all cases smaller than for model M_L and therefore confirmed the improvement of the results compared to model M_{LI}. Additionally, h_{het} was calculated for a model with larger parameter values, i.e., $\beta_0 = 2$, $\beta_1 = 2$, $\beta_2 = 1.2$, $\beta_5 = 1.2$ and $n = 400$. The power was high (99.99 %), and Type I error rate was only slightly increased (5.62 %).

Table 4 presents the results for the full nonlinear population model M_{LQI}. Type I error rates were again close to the nominal 5 % level, whereas three values were slightly too high. A power of 80 % was exceeded in samples with $n > 1,000$ when the nonlinear effect sizes were $\beta_3 = \beta_4 = .15$, $\beta_5 = .20$. Models with small effects required sample sizes of $n = 2,200$ in order to reach a power of 80 % (not listed in Table 4).

Heteroscedasticity unrelated to observed predictors

The following results relate to the investigation of heteroscedasticity due to unobserved predictors. The influence of nonlinear effect size and sample size on the h_{het}-values were examined. The results of the quadratic population model were compared to results of an analysis with the Breusch-Pagan test.

Linear population model. For the population model M_{sL} with a single linear effect the mean values of h_{het} and the Type I error rates are given in Table 5. The Type I error rates were close to the nominal 5 % level in all conditions ranging from 4.30 % to 5.62 %. Four Type I error rates were lying outside the 95 % CI bounds.

Nonlinear population models. The power of h_{het} to detect heteroscedasticity related to an unobserved predictor in a moderator model is provided in Table 6. M_{LI} was analyzed as M_{sL}, a model with a single linear predictor x_1. Even though the second linear predictor x_2 was not included in the analysis model, h_{het} responds to the heteroscedasticity caused by the interaction of the observed predictor x_1 and of the unobserved predictor x_2. The power was close to 80 % for $n = 800$ and $\beta_3 = .35$.

Table 7 reports the influence of an unobserved predictor in a quadratic population model on h_{het}. The predictor x_2 that was part of M_{LQ} was not included in the analysis model M_{sL} (a linear model with only a single linear predictor). The measure h_{het} responds to the heteroscedasticity associated to omitted predictors in M_{sL}. The power ranged from 16.31 % ($n = 100$, $\beta_3 = .15$) to 98.59 % ($n = 1,200$, $\beta_3 = .25$) and approximately exceeded 80 % for $n = 500$ and $\beta_3 = .25$ (not listed in Table 7). For reasons of comparability, the results for an analysis with the Breusch-Pagan test are listed too. The power values of the Breusch-Pagan test were lower and ranged from 7.98 % to 54.62 %. The reason why the Breusch-Pagan test does indeed have a power clearly above 5 % lies in the fact that x_1 and x_2 were correlated in model M_{LQ} for the simulated data.

Table 3: Mean h_{het}-Values, Type I Error Rates (in Percent), and Power (in Percent) as a Function of Sample Size (n) and Interaction Effect Size for the Population Model M_{LI}.

population model:	M_{LI} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_1 x_2$							
analysis model:	M_{LI} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_1 x_2$				M_L $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$			
nonlinear pop. parameter:	$\beta_5 = .30$		$\beta_5 = .40$		$\beta_5 = .30$		$\beta_5 = .40$	
n	Mean	Type I error	Mean	Type I error	Mean	Power	Mean	Power
100	-0.11	4.48	-0.12	4.40	0.06	7.85	0.41	13.62
200	-0.07	5.40	-0.09	5.49	0.33	12.37	0.93	25.09
400	-0.06	5.27	-0.05	5.49	0.59	17.37	1.61	40.18
800	-0.07	4.95	-0.04	5.76	1.04	28.15	2.57	63.45
1,200	-0.05	5.10	-0.04	5.41	1.32	36.24	3.23	77.22

Table 4: Mean h_{het}-Values, Type I Error Rates (in Percent), and Power (in Percent) as a Function of Sample Size (n) and Nonlinear Effect Size for the Full Nonlinear Population Model M_{LQI}.

population model:	M_{LQI} $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1^2+\beta_4x_2^2+\beta_5x_1x_2$							
analysis model:	M_{LQI} $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1^2+\beta_4x_2^2+\beta_5x_1x_2$				M_L $y=\beta_0+\beta_1x_1+\beta_2x_2$			
nonlinear pop. parameter:	$\beta_3=\beta_4=.10$ $\beta_5=.15$		$\beta_3=\beta_4=.15$ $\beta_5=.20$		$\beta_3=\beta_4=.10$ $\beta_5=.15$		$\beta_3=\beta_4=.15$ $\beta_5=.20$	
n	Mean	Type I error	Mean	Type I error	Mean	Power	Mean	Power
100	-0.09	4.96	-0.10	4.72	0.08	8.00	0.56	17.04
200	-0.07	4.77	-0.07	5.47	0.35	13.26	1.26	29.43
400	-0.06	5.41	-0.07	5.38	0.67	19.21	2.21	48.56
800	-0.06	5.21	-0.04	5.39	1.05	27.89	3.48	72.28
1,200	-0.05	5.44	-0.04	5.48	1.37	36.22	4.39	84.81

Table 5: Mean h_{het}-Values and Type I Error Rates (in Percent) as a Function of Sample Size (n) for the Population Model M_{sL} with a Single Linear Effect.

population model:	M_{sL} $y = \beta_0 + \beta_1 x_1$					
analysis model:	M_{sL} $y = \beta_0 + \beta_1 x_1$		M_{LQ} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_2^2$		M_{LI} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ $\beta_5 x_1 x_2$	
n	Mean	Type I error	Mean	Type I error	Mean	Type I error
100	-0.13	4.30	-0.11	4.69	-0.12	4.75
200	-0.08	5.36	-0.08	5.36	-0.10	5.12
400	-0.07	5.36	-0.06	5.17	-0.04	5.62
800	-0.04	5.23	-0.04	5.57	-0.05	4.98
1,200	-0.03	5.45	-0.04	5.33	-0.03	5.13

Table 6: Mean h_{het}-Values and Power (in Percent) as a Function of Sample Size (n) and Interaction Effect Size for the Population Model M_{LI}.

population model:	M_{LI} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_1 x_2$			
analysis model:	M_{sL} $y = \beta_0 + \beta_1 x_1$			
nonlinear pop. parameter:	$\beta_5 = .25$		$\beta_5 = .35$	
n	Mean	Power	Mean	Power
100	0.32	12.90	0.68	19.90
200	0.63	18.60	1.52	36.10
400	1.13	31.80	2.14	52.00
800	1.64	43.10	3.24	76.20
1,200	1.92	53.00	3.92	87.30

Under the same conditions, except for using uncorrelated predictor terms, the power of the Breusch-Pagan test dropped below 14 %, while the power of h_{het} was unaffected (not listed in Table 7). A comparison with the White test (not reported here) revealed that the White test had slightly lower power than the Breusch-Pagan test. Therefore, only results for the Breusch-Pagan test are reported here. The power of the White test ranged from 5.85 % to 36.5 %.

Table 7: Mean h_{het}- and Breusch-Pagan-Values and Power (in Percent) as a Function of Sample Size (n) and Quadratic Effect Size for the Population Model M_{LQ}.

heteroscedasticity analysis:	h_{het}				*Breusch-Pagan test*[1]			
population model:	M_{LQ} $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_2^2$							
analysis model:	M_{sL} $y = \beta_0 + \beta_1 x_1$							
nonlinear pop. parameter:	$\beta_4 = .15$		$\beta_4 = .25$		$\beta_4 = .15$		$\beta_4 = .25$	
n	Mean	Power	Mean	Power	Mean	Power	Mean	Power
100	0.55	16.31	1.58	33.23	0.46	7.98	0.42	12.58
200	0.96	24.98	2.73	52.58	0.44	10.61	0.36	19.08
400	1.49	36.57	4.31	75.95	0.39	14.18	0.30	26.71
800	2.24	54.84	6.41	93.96	0.31	22.84	0.21	42.62
1,200	2.82	69.15	7.96	98.59	0.27	30.63	0.15	54.62

[1] The formula for the Breusch-Pagan test was $\frac{e^2}{\tilde{\sigma}^2} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \varepsilon$, where $\tilde{\sigma}^2 = \frac{\Sigma e_i^2}{n}$ and ε is normally distributed with zero mean.

Empirical example

To illustrate the applicability of h_{het} an empirical example is presented where the influence of job characteristics on burnout was examined. The dependent variable is emotional exhaustion, which is considered to be one central symptom of the burnout syndrome (Maslach & Jackson, 1981; Maslach & Leiter, 1997). Exhaustion refers to feelings of being overextended and drained by job demands. Three predictors were considered: Job control, work pressure, and concentration requirements. Work pressure involves perceived time pressure and work volume, and concentration requirements refer to employee's experienced degree of task complexity and demands on concentration.

Participants and procedure. The study was carried out in a large civil service organization of a federal state in Germany (Diestel & Schmidt, 2009; Schmidt & Neubach, 2009). Participants of the study were tax collectors, recruited from a large tax and revenue office. During work hours, questionnaires were administered to 641 employees in small groups of about 15 people. A final sample of 461 employees provided sufficient data. Mean age was 40.88 (SD = 10.05), 58 % of the employees were female and 89.6 % were employed on a full-time basis.

Measures. The burnout dimension of emotional exhaustion was measured by Büssing and Perrar's (1992) German translation of the Maslach Burnout Inventory (Maslach, Jackson, & Leiter, 1986). Nine items measured emotional exhaustion (e.g., 'I feel emotionally drained from my work'). Job control was measured by five items, which refer to the perceived extent to which an employee can choose different strategies and methods (Jackson, Wall, Martin, & Davids, 1993; Schmidt, 2004) (e.g., 'To what extent can you decide how to go about getting your job done?'). Work pressure and concentration requirements, two dimensions of work load, were measured by subscales of the Kurzfragebogen zur Arbeitsanalyse (KFZA; Short Questionnaire for Job Analysis) instrument developed by Prümper, Hartmannsgruber, and Frese (1995). Both scales, originally measured by two items each, were extended by constructing two additional items for work pressure and three additional items for concentration requirements (Schmidt & Neubach, 2009).

Results. Two regression analyses were conducted. First, a linear model was analyzed, where 'emotional exhaustion' (EE) was regressed on 'work pressure' (WP), 'concentration requirements' (CR), and 'job control' (JC):

$$EE = \beta_0 + \beta_1 WP + \beta_2 CR + \beta_3 JC + e \tag{15}$$

The OLS regression for the linear model yielded $R^2 = .47$, the standardized regression equation is:

$$\hat{z}_{EE} = .27 z_{WP} + .32 z_{CR} - .26 z_{JC}. \tag{16}$$

All three linear effects were significant (with $t = 5.93$, $SE = .045$, $p < .01$ for predictor WP; $t = -6.95$, $SE = 0.038$, $p < .01$ for predictor JC; $t = 7.13$, $SE = .045$, p < .01 for predictor CR). The analysis of the residuals resulted in $h_{het} = 1.84$ for the linear model. For $\alpha = 5\%$ the critical h_{het} value is 1.65. Thus, the residuals showed significant heteroscedasticity in the linear regression model. It can be inferred that possible moderator and nonlinear effects may have been omitted in the linear model. The AIC for this model was 1019.82.

In order to detect the origin of the heteroscedasticity, a second regression model with multiple nonlinear effects was analyzed. As job control is expected to buffer the positive effect of work pressure on emotional exhaustion (cf. Häusser, Mojzisch, Niesel, & Schulz-Hardt, 2010; Karasek, 1979) the interaction effect of work pressure and job control $(WP \times JC)$ was included in the regression equation. Additionally, quadratic terms were included for the predictor WP and JC, because this can reduce the risk of a spurious interaction (cf. Cortina, 1993; Klein, Schermelleh-Engel, Moosbrugger, & Kelava, 2009).

$$EE = \beta_0 + \beta_1 WP + \beta_2 CR + \beta_3 JC + \beta_4 WP^2 + \beta_5 JC^2 + \beta_6 WP \times JC + e \qquad (17)$$

Before forming product variables, we standardized the predictor variables in order to reduce multicollinearity and to obtain a correctly standardized solution (Aiken & West, 1991). The OLS regression for the nonlinear model yielded $R^2 = .51$, the standardized regression equation is:

$$\hat{z}_{EE} = -.07 + .26 z_{WP} + .29 z_{CR} - .29 z_{JC} + 0.07 z_{WP^2} - 0.06 z_{JC^2} - .14 z_{WP} \times z_{JC}. \qquad (18)$$

Besides significant linear effects, the quadratic effect of work pressure (with $t = 2.48$, $SE = .027$, $p = .01$) and the interaction effect (with $t = -4.2$, $SE = .034$, $p < .01$) were significant, while the quadratic effect of job control (with $t = -1.88$, $SE = .029$, $p = .06$) just failed to reach statistical significance. Compared to the linear model the value of h_{het} was reduced to $h_{het} = 1.14$. As this value was smaller than the critical value (1.65) it was concluded that the residuals were now homoscedastic in the nonlinear model. The nonlinear terms explained satisfactorily all the heteroscedasticity that appeared in the residuals of the linear model. The AIC value of 996.08 also indicated an improved model fit compared to the linear model ($AIC = 1019.82$).

Figure 3 gives estimated histograms of h_{het} for both linear (left panel) and nonlinear models (right panel). The h_{het}-values were estimated using bootstrapping with 10,000 replications. According to our expectations, the distribution of the resampled h_{het}-values was shifted to the left when a nonlinear model was fit to the data. The kurtosis was -.05 for the linear model and .08 for the nonlinear model. The linear model had a skewness of .15, the nonlinear model a skewness of .32.

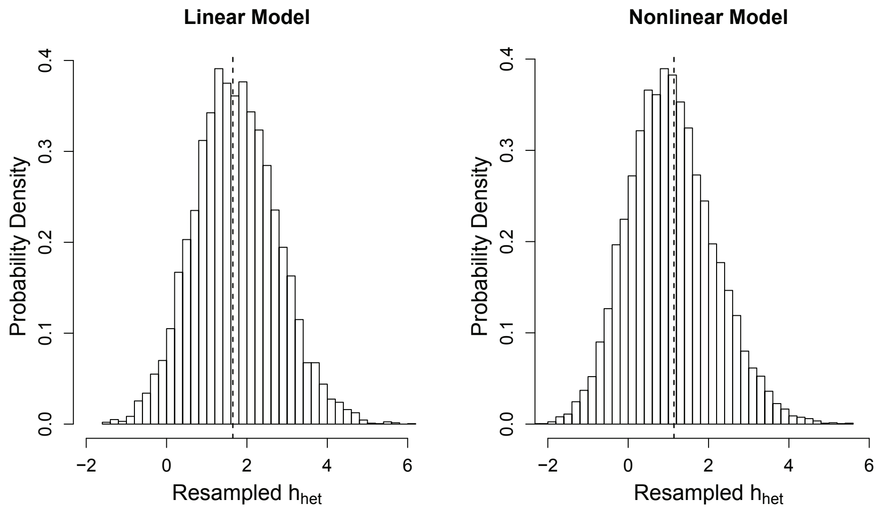

Figure 3: Bootstrapped h_{het}-values for the linear and nonlinear model of emotional exhaustion. 10,000 data sets were resampled.

Our results of the empirical study are well in line with the Job Demands-Resources (JD-R) model (Bakker, Demerouti, De Boer, & Schaufeli, 2003; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001), a model often used to explain how job strain (e.g., burnout) may be produced by two working conditions, for example, job demands and job resources (see also Diestel & Schmidt, 2009). The results revealed a buffering effect of job control on the relationship between work pressure and emotional exhaustion: For high values of job control, the enhancing effect of work pressure on emotional exhaustion is diminished. Additionally, we found a quadratic effect (WP^2). While this effect was not particularly large, it indicates that the effect of quantitative work stress on burnout is especially severe under high levels of work pressure.

Discussion

In this article, we proposed the measure h_{het} for detecting heteroscedasticity in regression analysis. This measure utilizes the kurtosis of the residuals in a new context and makes direct use of the dispersion of the squared residuals. In contrast to other heteroscedasticity tests (e.g., Breusch & Pagan, 1979; White, 1980), it does not require a specific parameterization of heteroscedasticity.

In a Monte-Carlo Study we tested the performance of h_{het}. The results indicate the ability of the measure to respond to model misspecification caused by nonlinear predictor terms omitted in the analyzed model. A power analysis demonstrated the need of sufficiently large sample size when small nonlinear effects are omitted. We did not investigate the performance of h_{het} for particularly small sample sizes. It is evident from our results that the statistical power would be too low in this case. A Type I error analysis showed encouraging results, the Type I error rate was never higher than 5.76 % and therefore only slightly increased. Thus, the measure h_{het} could be used in regression analysis to identify heteroscedastic errors. For one simulation condition, h_{het} was compared to the *AIC*. The *AIC* showed the necessity of the nonlinear terms in all simulated datasets. Still, it should be noted that the *AIC* cannot be used to detect heteroscedasticity related to unobserved predictors. For heteroscedasticity due to omitted predictors, the power of h_{het} was considerably higher than the power of the Breusch-Pagan test. This was expected, because the Breusch-Pagan test only detects explanatory variables that are related to the error variances (Breusch & Pagan, 1979). On the other hand, if heteroscedasticity is caused by the observed predictors, residual based tests such as the Breusch-Pagan test are preferable. Still, h_{het} does also respond to this kind of heteroscedasticity, but with lower power. The applicability of h_{het} was further demonstrated by an empirical example from psychology, where a regression model with linear terms was shown to have heteroscedastic error terms related to omitted nonlinear terms. The h_{het}-value responded to the fact that the model was misspecified when nonlinear predictor terms were omitted.

Regression models have been used in the social sciences at least since 1899, when Yule published a paper on the causes of pauperism (Yule, 1899). At present, regression models are state-of-the-art not only for the social and behavioral sciences, but also across scientific disciplines. In order to enhance prediction, nonlinear effects, i.e. interaction and quadratic effects, have been added to the linear regression equation. The use of interaction effects has increased significantly since Aiken and West's (1991) seminal book on moderated regression. In psychological research overlooked or yet unidentified moderator variables go typically along with omitted product terms in regression. Adding an interaction term to a regression model can therefore greatly enhance the understanding of the relationships among the variables in the model. For example, in the context of burnout research, several studies have demonstrated buffering effects of diverse resources on the relationship between stress and strain (cf. Gray-Stanley & Muramatsu, 2011; Schmidt, 2007). Additionally, curvilinear effects on burnout have been found, for example, between work ambiguity on burnout (Jamal, 2008) and between job demands and anxiety (de Jonge & Schaufeli, 1998).

The present study has some important limitations. First, we examined the measure under ideal distributional conditions where the residuals were all normally distributed. Future simulation studies are needed to test the robustness of h_{het} to violations of the normality assumption. Second, the effect of strong overparameterization should be investigated in a simulation study. In practice, the researcher should pay attention to the fact that a strongly overparameterized model can lead to wider confidence intervals.

One should keep in mind that the h_{het} measure is not constructive, which means that a significant h_{het}-value provides no specific information about the source of the heteroscedasticity in the data. There may exist different possible reasons for heteroscedasticity in multiple regression. One possible reason is the presence of outliers in the data, which should be checked routinely before performing regression analysis and before applying the measure h_{het}. In multiple regression an incorrectly specified regression model, where important variables are omitted or where the functional form is incorrect, may produce significant results when testing heteroscedasticity. In order to analyze this type of heteroscedasticity, the Breusch-Pagan test is well suited when the predictors that form the nonlinear terms are observed. The new measure h_{het} is advantageous and could be used if nonlinear terms of unknown predictor variables are assumed to having been omitted in the study. For this purpose, theoretical considerations about possible model misspecifications and other potential sources of heteroscedastic residuals are necessary.

Author note

This research was supported by Grant No. SCHE1412-1/1 from the German Research Foundation (DFG).

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Bakker, A. B., Demerouti, E., De Boer, E., & Schaufeli, W. B. (2003). Job demands and job resources as predictors of absence duration and frequency. *Journal of Vacational Behavior*, *62*(2), 341–356. doi: 10.1016/S0001-8791(02)00030-1

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*(5), 1287–1294. doi: 10.2307/1911963

Büssing, A., & Perrar, K.-M. (1992). Die Messung von Burnout. Untersuchung einer deutschen Fassung des Maslach Burnout Inventory (MBI-D)[Burnout measurement. A Study of a German version of the Maslach Burnout Inventory (MBI-D)]. *Diagnostica*, *38*, 328–353.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, *70*(1), 1–10. doi: 10.2307/2335938

Cortina, J. M. (1993). Interaction, nonlinearity and multicollinearity: Implications for multiple regression. *Journal of Management*, *19*(4), 915 – 922.

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.

de Jonge, J., & Schaufeli, W. B. (1998). Job characteristics and employee well-being: A test of Warr's vitamin model in health care workers using structural equation modelling. *Journal of Organizational Behavior*, *19*(4), 387–407.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499–512.

Diestel, S., & Schmidt, K.-H. (2009). Mediator and moderator effects of demands on self-control in the relationship between work load and indicators of job strain. *Work & Stress*, *23*(1), 60–79. doi: 10.1080/02678370902846686

Dijkstra, T. K., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, *79*(4), 585–604. doi: 10.1007/s11336-013-9370-0

Engle, R. F. (1984). Wald, likelihood ratio and Lagrange multiplier tests in econometrics. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of Econometrics* (Vol. 2, pp. 775–826). Amsterdam: Elsevier.

Gray-Stanley, J. A., & Muramatsu, N. (2011). Work stress, burnout, and social and personal resources among direct care workers. *Research in Developmental Disabilities*, *32*(3), 1065–1074. doi: 10.1016/j.ridd.2011.01.025

Greene, W. H. (2003). *Econometric Analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Häusser, J. A., Mojzisch, A., Niesel, M., & Schulz-Hardt, S. (2010). Ten years on: A review of recent research on the job demand-control (-support) model and psychological well-being. *Work & Stress*, *24*(1), 1–35. doi: 10.1080/02678371003683747

Jackson, P. R., Wall, T. D., Martin, R., & Davids, K. (1993). New measures of job control, cognitive demand, and production responsibility. *Journal of Applied Psychology*, *78*(5), 753–762. doi: 10.1037/0021-9010.78.5.753

Jamal, M. (2008). Burnout among employees of a multinational corporation in Malaysia and Pakistan: An empirical examination. *International Management Review*, *4*(1), 60–71.

Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, *24*(2), 285–308. doi: 10.2307/2392498

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*(4), 457–474. doi: 10.1007/BF02296338

Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*(4), 647–673. doi: 10.1080/00273170701710205

Klein, A. G., & Schermelleh-Engel, K. (2010). Introduction of a new measure for detecting poor fit due to omitted nonlinear terms in SEM. *AStA Advances in Statistical Analysis*, *94*(2), 157-166. doi: 10.1007/s10182-010-0130-5

Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Kelava, A. (2009). Assessing spurious interaction effects. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (p. 13-28). Rotterdam, NL: Sense Publishers.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305-325. doi: 10.1016/0304-4076(85)90158-7

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behaviour*, *2*(2), 99-113. doi: 10.1002/job.4030020205

Maslach, C., Jackson, S. E., & Leiter, M. P. (1986). *Maslach Burnout Inventory* (2nd ed.). Palo Alto, CA: Consulting Psychologist Press.

Maslach, C., & Leiter, M. P. (1997). *The Truth About Burnout*. San Francisco, CA: Jossey-Bass.

Prümper, J., Hartmannsgruber, K., & Frese, M. (1995). KFZA. Kurz-Fragebogen zur Arbeitsanalyse [Short-Questionnaire for Job Analysis]. *Zeitschrift für Arbeits- und Organisationspsychologie*, *39*(3), 125-132.

R Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.2.) [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing.

Rosopa, P. J., Schaffer, M. M., & Schroeder, A. M. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods*, *18*(3), 335-351. doi: 10.1037/a0032553

Schmidt, K.-H. (2004). Formen der Kontrolle als Puffer der Belastungs-Beanspruchungs-Beziehung [Forms of control as buffer of the relationship between job demands and strain]. *Zeitschrift für Arbeitswissenschaft*, *58*, 44-52.

Schmidt, K.-H. (2007). Organizational commitment: A further moderator in the relationship between work-stress and strain? *International Journal of Stress Management*, *14*(4), 26–40. doi: 10.1037/1072-5245.14.1.26

Schmidt, K.-H., & Neubach, B. (2009). Selbstkontrollanforderungen als spezifische Belastungsquelle bei der Arbeit [Self-control demands as a specific source of stress at work]. *Zeitschrift für Personalpsychologie*, *8*(4), 169-179. doi: 10.1026/1617-6391.8.4.169

Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, *95*(2), 334-344. doi: 10.1037/0033-2909.95.2.334

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*(3), 426-482.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838. doi: 10.2307/1912934

Yule, G. U. (1899). An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades (part I.). *Journal of the Royal Statistical Society*, *62*(2), 249–295. doi: 10.2307/2979889