



Investigating the quantity–quality relationship in scientific creativity: an empirical examination of expected residual variance and the tilted funnel hypothesis

Boris Forthmann¹ · Mark Leveling² · Yixiao Dong² · Denis Dumas²

Received: 4 March 2020 / Published online: 25 June 2020
© The Author(s) 2020

Abstract

Among scientists who study scientific production, the relationship between the quantity of a scientist’s production and the quality of their work has long been a topic of empirical research and theoretical debate. One principal theoretical perspective on the quantity–quality relationship has been the equal odds baseline, which posits that a scientist’s number of high-quality products increases linearly with their total number of products, and that there is a zero correlation between a scientist’s total number of products and the average quality of those products. While these central tenets of the equal odds baseline are well known, it also posits a number of more specific and less discussed aspects of the quality–quantity relation, including the expected residual variance and heteroscedastic errors when quality is regressed on quantity. After a careful examination of the expected variance by means of a non-parametric bootstrap approach, we forward a further prediction based on the heteroscedasticity implied by the equal-odds baseline that we term the *tilted funnel hypothesis*, that describes the shape of a bivariate scatterplot when quality is regressed on quantity, as well as the change in the strength of slope coefficients at different conditional quantiles of the quality distribution. In this study, we empirically test the expected residual variance and the tilted funnel hypothesis across three large datasets (including approximately 1.5 million inventors, 1800 psychologists, and 20,000 multidisciplinary scientists). Across all of the data sets, the results empirically supported the tilted funnel hypothesis, and therefore the results provided further evidence of the utility of the equal odds baseline.

Keywords Quantity · Quality · Publications · Patents · Equal odds baseline · Heteroscedasticity · Residual variance · Quantile regression

Mathematics Subject Classification 62J02

JEL Classification C31

✉ Boris Forthmann
boris.forthmann@wwu.de

¹ Institute of Psychology in Education, University of Münster, Münster, Germany

² Department of Research Methods and Information Science, Morgridge College of Education, University of Denver, Denver, CO, USA

Introduction

The interplay of quantity of scientific productions (e.g., publications or patents) and the perceived quality of those scientific productions in the field (e.g., the number of received citations) is one of the most fundamental issues in research on scientific creativity (Feist 1997; Lawani 1986; Simonton 1988). In general, the quantity–quality relationship has been studied to better understand the way that scientists’ productivity, or quantity of publications, may affect the influence that those publications have in the field—typically operationalized as citations—with researchers in this area at times uncovering a potential trade-off between quality and quantity (Fischer et al. 2012), and at times identifying a trend in which productivity supports the quality of publications (often depending on the used operational definition of quality; see Michalska-Smith and Allesina 2017).

A classical theoretical model that has been perennially used to describe this relation is Simonton’s equal odds baseline in which the number of high-quality publications is considered to be a linear function of total publications (e.g., Simonton 1988, 2004, 2010). The equal odds baseline is a parsimonious model for the relation between quantity and quality of scientific productions that is based on several assumptions. For example, the equal odds baseline posits that scientists’ total publication output is uncorrelated with the average quality of those publications: a theoretical prediction that is testable empirically. Relatedly, Forthmann et al. (2020a) demonstrated how the equal-odds baseline can be derived from the assumption that average quality and total output are independent and introduced a structural equation modeling approach to assess the fit of the equal odds baseline for a given dataset by means of widely used fit indices and established cut-offs. In this work, we aim at extending the knowledge base related to the equal odds baseline by focusing on two aspects of residual variance. First, the expected residual variance under the equal odds baseline will be investigated. Second, it has been argued that the equal odds baseline applies on average in the population (Simonton 2003a, b, 2004, 2009) which fits conceptually well with correlational and conditional mean regression approaches. However, the equal odds baseline also predicts a skewed distribution of residual terms that exhibit dependence on quantity (e.g., Simonton 2010). Clearly, such residuals imply heteroscedastic errors (i.e., non-constant variance across the values of quantity) in a regression context and, as a consequence, the question emerges if the proposed linear relationship between quantity and quality generalizes to the full conditional distribution of publication quality. Hence, the aim of this work is to thoroughly investigate the relationship between quantity and quality by means of quantile regression (Koenker 2005; Wenz 2019; Yu et al. 2003) which elegantly extends the focus on the conditional mean to the full conditional distribution of publication quality.

The equal odds baseline

The equal odds baseline (EOB) refers to the linear regression of the conditional mean of the number of hits H (i.e., the number of high-quality publications) on the total number of productions T (i.e., the total number of publications; see Simonton 1988, 2003a, 2004, 2009, 2010). The prediction of the conditional mean corresponds well with Simonton’s assertion that the EOB theoretically works for the population average (Simonton 2003a, b, 2004, 2009). A random shock term u is further added to the model to take individual deviations from the average hit rate ρ into account (Cole and Cole 1967; Feist 1997):

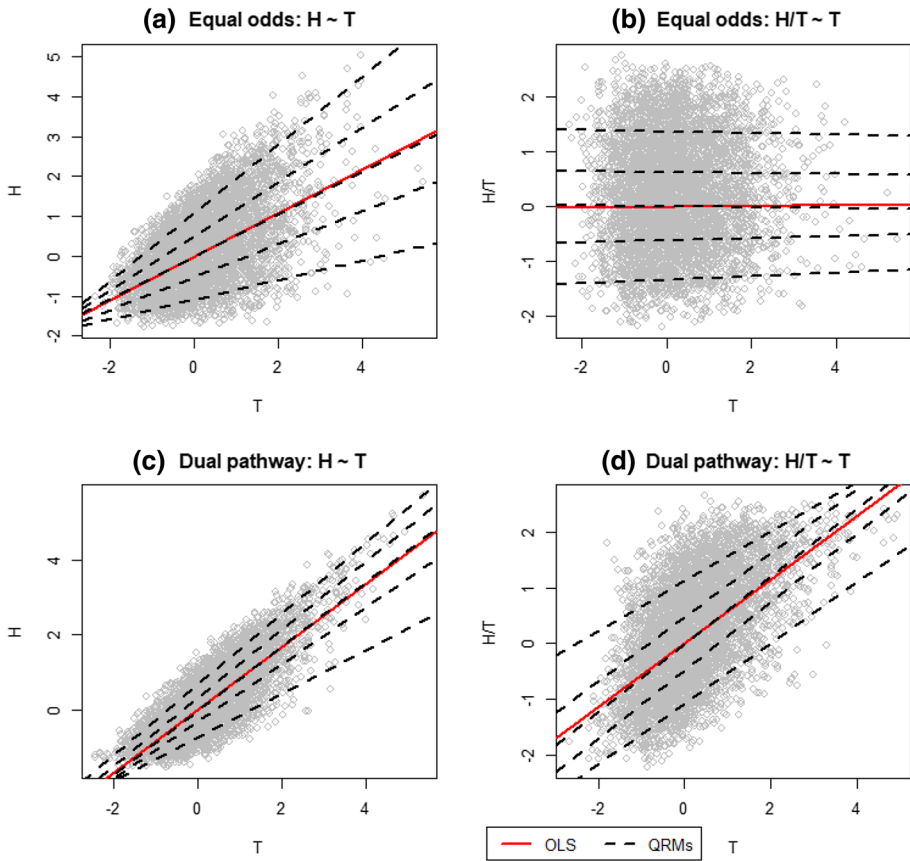


Fig. 1 Bivariate Scatterplots for the relationship between z -standardized H (y-axis) and z -standardized T (x-axis) or z -standardized H/T (y-axis) and z -standardized T (x-axis). Subplots (a) and (b) are based on simulated data adhering to the EOB. Subplots (c) and (d) are based on simulated data adhering to the dual pathway theory of creativity (i.e., the EOB is violated). OLS=regression slope from ordinary least squares regression (predicting the conditional mean). QRM5s=regression slopes from quantile regression to predict the .10, .30, .50, .70, and .90 quantiles of the conditional distribution. The R code to reproduce these plots are available in the Open Science Framework (<https://osf.io/2zgxn/>)

$$H = \rho T + u. \tag{1}$$

Equation 1 refers to the cross-sectional EOB. The EOB is illustrated by the red regression slope in Fig. 1a. Altogether $N=5000$ data points adhering to the EOB were simulated for illustration purposes.¹ This model further assumes that the hit-ratio H/T is uncorrelated with T (see also the red regression line in Fig. 1b) because a negative or positive correlation between these two variables implies a non-linear (i.e., quadratic)

¹ All data in Fig. 1 were simulated according to a simulation setup used by Forthmann et al. (2020a, b). The exact parameters used in the simulation are available in the R script used to create Fig. 1. The R script is openly available in the Open Science Framework (OSF; <https://osf.io/2zgxn/>).

relation between H and T (Simonton 2003a, b, 2004). Hence, a test of the correlation between H/T and T is sensitive for deviations from the EOB (see Forthmann et al. 2019).

In addition, Forthmann et al. (2020a) have shown that ρ equals $E(H)/E(T)$ when the EOB is derived from the assumption that H/T and T are independent. Importantly, H and T must be positive reals such that $E(H)$, $E(T)$, and $E(H/T)$ exist (see Forthmann et al. 2020b), but no further assumptions (e.g., with respect to the distribution of T and H or the upper limit of ρ) are required and, thus, their derivation is very general. Forthmann et al. (2020a) further outlined an approach how the EOB can be tested by means of structural equation modeling (SEM). The advantage of that approach is the availability of common fit measures from the extensive SEM literature. Hence, the fit of the EOB to a given dataset—i.e., the practical value of the EOB—can be assessed by well-established model fit thresholds for various indices (e.g., West et al. 2012). Another important consequence from $\rho = E(H)/E(T)$ is obtained for the residual variance. Inserting $\rho = E(H)/E(T)$ into the EOB, and taking into account that u is assumed to be uncorrelated with H and T , yields the following expression for the variance of H :

$$\text{Var}(H) = \rho^2 \text{Var}(T) + \text{Var}(u) = \left(\frac{E(H)}{E(T)} \right)^2 \text{Var}(T) + \text{Var}(u). \quad (2)$$

Hence, from Eq. 2 the expected residual variance when the EOB holds is obtained as

$$\text{Var}(u) = \text{Var}(H) - \left(\frac{E(H)}{E(T)} \right)^2 \text{Var}(T). \quad (3)$$

Importantly, the part in Eq. 3 relating to the linear predictor helps to understand deviations from the EOB. It follows that residual variance will be larger than expected under the EOB, when the regression slope is smaller as compared to $E(H)/E(T)$. This is, for example, the case when the correlation between H/T and T is negative (Forthmann et al. 2020a, b). In the creativity research literature, a negative correlation between average quality and quantity is in accordance with a trade-off between quality and quantity (i.e., high-quality ideas require more cognitive effort and, hence, more time; e.g., Forthmann et al. 2020b; Guilford 1968) which is an idea that is also discussed and studied in relation to scientific productions (e.g., Fischer et al. 2012; Michalska-Smith and Allesina 2017). Analogously, residual variance will be smaller than expected under the EOB, when the regression slope is larger as compared to $E(H)/E(T)$. This is, for example, the case when data adhere to the dual pathway theory implying a positive correlation between H/T and T (Forthmann et al. 2020a, b). The dual pathway theory of creativity (e.g., Nijstad et al. 2010) proposes two routes to achieve more original ideas on average: (a) The first route is via flexibility which refers to the number of conceptual categories from which ideas were generated (i.e., variety as a measure of diversity; see Stirling 2007), and (b) the second route is via persistence which refers to ideational exhaustion of each of the conceptual categories (i.e., the average number of ideas within each of the conceptual categories). Both routes imply both higher total ideas (i.e., T) and higher average originality (i.e., H/T) and, hence, the dual pathway theory implies that H/T is positively predicted by T (Forthmann et al. 2019, a, b; Nijstad et al. 2010; see also Fig. 1d).

Replacing expected values and variances of the random variables by sample estimators in Eq. 2 allows constructing a test statistic $\Delta_{\hat{\sigma}_u^2}$ for the difference between the observed and the expected residual variance under the EOB

$$\Delta_{\hat{\sigma}_u^2} = s_u^2 - \hat{\sigma}_u^2 = s_u^2 - s_H^2 - \left(\frac{\bar{H}}{\bar{T}}\right)^2 s_T^2, \tag{4}$$

with s_u^2 being an estimate of the observed residual variance, $\hat{\sigma}_u^2$ being the expected variance under the EOB as quantified based on the sample means and variances of quality and quantity, \bar{H} , \bar{T} , s_H^2 , and s_T^2 , respectively.

Influencing factors on the $\Delta_{\hat{\sigma}_u^2}$ statistic

Analogous to the EOB illustration in Fig. 1, we used simple simulations to illustrate how the correlation between H/T and T (levels of absolute r values: .11, .29, .47, and .64), the average of T (20 vs. 50), and the average ρ (i.e., average H/T ; levels: .22, .35, .5, .65, and .78 for ρ in the range from 0 to 1) influence the $\Delta_{\hat{\sigma}_u^2}$ statistic (see Fig. 2). Each point in Fig. 2 represents the $\Delta_{\hat{\sigma}_u^2}$ statistic based on 5000 simulated cases. The simulation setup is described in detail in Forthmann et al. (2020b). The full R script to run these illustratory simulations is available in the OSF repository (<https://osf.io/2zgxn/>). It should be noted that for the sake of simplicity we do not distinguish in this work between random variables and realizations thereof by means of the used notation. When required the respective meaning is made salient to the reader.

Expectedly, the size of the correlation between H/T and T influenced the deviation of the observed residual variance from its expected value under the EOB (see Fig. 2). The $\Delta_{\hat{\sigma}_u^2}$ statistic is positive and increases away from zero with stronger negative correlations between H/T and T (see top plots in Fig. 2), whereas negative and decreasing values away from zero were observed as positive correlations between H/T and T increase in strength (see bottom plots in Fig. 2). As noted above, this behavior of the $\Delta_{\hat{\sigma}_u^2}$ statistic depends on the variance attributable to the linear predictor that is either smaller (e.g., when the correlation between H/T and T is negative) or larger (e.g., when the correlation between H/T and T is positive) as compared to the expected value under the EOB. In addition, the $\Delta_{\hat{\sigma}_u^2}$ statistic deviates more from zero as the average T increases (compare the left and right plots in Fig. 2).

The pattern for $\Delta_{\hat{\sigma}_u^2}$ as a function of average ρ , however, seems to be more complex. For negative correlations between H/T and T and ρ ranging between 0 and 1 (see the top plot in Fig. 2), there is an interaction between the size of the correlation between H/T and T and average ρ . Specifically, the differences in $\Delta_{\hat{\sigma}_u^2}$ values between the levels of average ρ increase with the strength of the correlation between H/T and T (this pattern is most obvious in the top-right plot in Fig. 2), resulting in larger $\Delta_{\hat{\sigma}_u^2}$ values for higher average ρ . Importantly, for rather low values of average ρ (e.g., .22) the $\Delta_{\hat{\sigma}_u^2}$ statistic can yield comparable values for small ($r = -.11$) and large ($r = .64$) deviations from the EOB when ρ ranges from 0 to 1 (see top-right plot in Fig. 2). Clearly, the lower the average ρ , the smaller the range of hit-ratios H/T (i.e., average ρ approaches its lower bound of zero) which in turn can limit the likelihood that negative correlations between H/T and T unfold in empirical data. This pattern is reversed for positive correlations between H/T and T (see bottom plots in Fig. 2). That is, $\Delta_{\hat{\sigma}_u^2}$ decrease away from zero as a function of both the strength of the correlation between H/T and T and average ρ . However, when average ρ approaches its upper limit of 1, $\Delta_{\hat{\sigma}_u^2}$ values become more comparable to each other (e.g., for average ρ of .50 and .65; see bottom-right plot in Fig. 2) and, finally, seem to turn back to zero when average ρ further approaches 1 (e.g., $\Delta_{\hat{\sigma}_u^2}$ is closer to zero for an average ρ of .78 as compared to average ρ of .35 for the highest simulated correlation between H/T and T ; see bottom-right plot in Fig. 2).

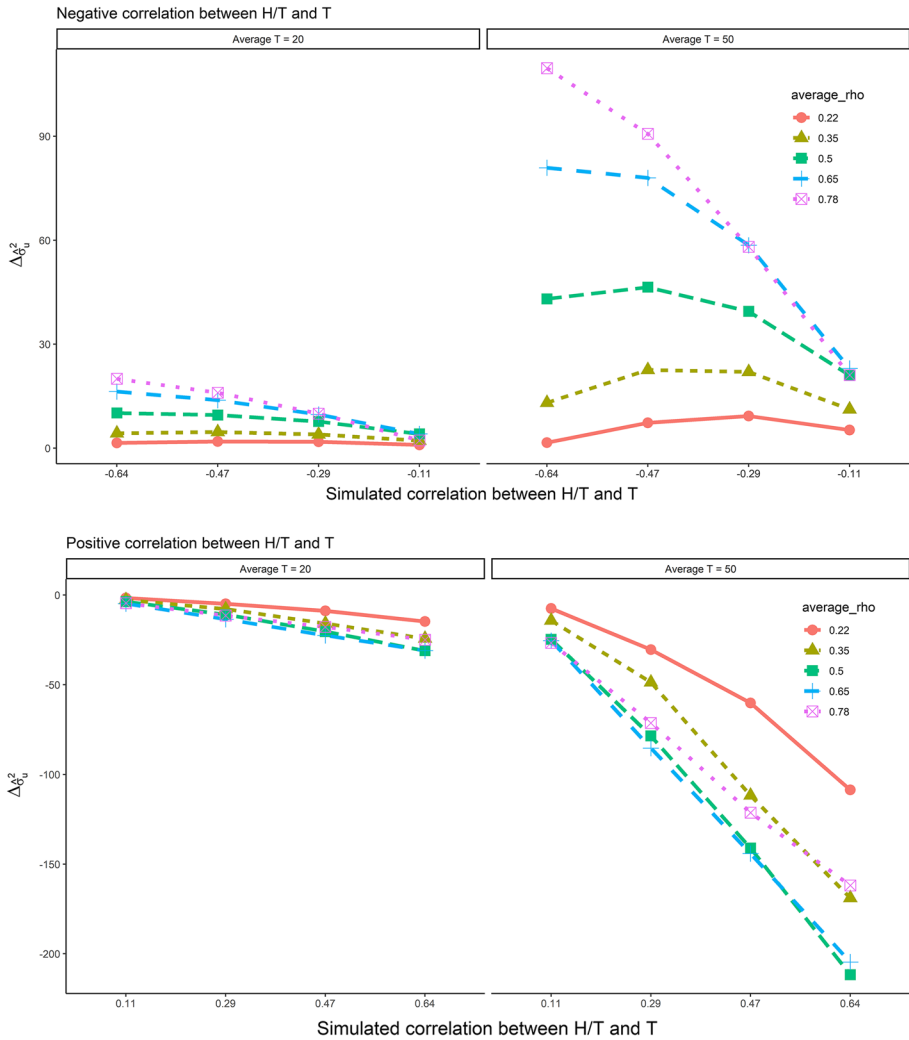


Fig. 2 The difference between the observed and expected residual variance is denoted by $\Delta\sigma_u^2$ and depicted on the y-axes for ρ values in the range of 0–1 (i.e., as in the original formulation of the equal-odds baseline). The simulated correlation between H/T and T is depicted on the x-axes. Each plot includes different lines and symbols according to the level of the simulated average ρ (i.e., average H/T ; see plot legends). *Top* negative correlations between H/T and T . *Bottom* positive correlations between H/T and T . *Left* average $T=20$. *Right* average $T=50$. The R code to reproduce this figure is available in the Open Science Framework (<https://osf.io/2zgxn/>)

Importantly, the overall illustration in Fig. 2 does not change when H/T is scaled by a factor of eight (i.e., ρ is not bounded above by 1) in the simulation (see additional material in the OSF: <https://osf.io/2zgxn/>). This suggests that distributional properties of H/T and the conditional distribution $f(H/T)$ help to explain the behavior of the $\Delta\sigma_u^2$ statistic more generally. Indeed, the simulated skewness of H/T varies as a function of average ρ with the highest positive skew observed for the lowest average ρ (i.e., average $\rho = .22$; skew ≈ 1.00)

over the mid-level of both skew and ρ (i.e., average $\rho = .50$; skew $\approx .00$) to the highest level of average ρ being associated with the lowest negative skew (i.e., average $\rho = .78$; skew ≈ -1.00). A highly comparable pattern of average skewness coefficients as a function of average ρ results for the simulated conditional distributions $f(H/T)$. In relation to this, it should further be mentioned that negative correlations between H/T and T tend to increase skewness, whereas positive correlations between H/T and T tend to decrease skewness. Hence, for ρ in general it can be concluded that deviations from the EOB as indicated by positive values of the $\Delta_{\delta_u^2}$ statistic (e.g., in the case of negative correlations between H/T and T) would be more pronounced the stronger the correlations and for negatively skewed distributions of H/T and $f(H/T)$ as compared to positively skewed distributions of H/T and $f(H/T)$. Likewise, for positive correlations between H/T and T , the size of the correlations in combination with the skewness of the distribution of H/T and the conditional distribution $f(H/T)$ also matters: stronger correlations in combination with positively skewed over symmetric to slightly negatively skewed distributions of H/T and $f(H/T)$ would be associated with decreasing values of the $\Delta_{\delta_u^2}$ statistic. However, when negative skewness coefficients for the distribution of H/T (and $f(H/T)$) further decrease, this is expected to result into a shifting back to zero for the $\Delta_{\delta_u^2}$ statistic.

Finally, the overall distribution of quality $f(H)$ can be positively skewed even when all conditional distributions $f(H/T)$ are negatively skewed because of the dependence between H and T . In fact, including negatively skewed $f(H/T)$ and $f(H)$ in the illustration above insured a more comprehensive treatment of the phenomenon. However, distributions of citation counts are commonly found to be positively skewed and simulated as such in the bibliometric literature (e.g., Blagus et al. 2015; Saam and Reiter 1999). In addition, also average citations per publication (i.e., the impact factor) commonly tend to be positively skewed. For example, Calver and Bradley (2009) found average citation per publication to be positively skewed for all journals (varying in number of publications over the period from 2000 through 2006) from which articles were cited. Hence, for empirical data, the distribution of H/T and most of the conditional distributions $f(H/T)$ are expected to be positively skewed. Likewise, scientific productions of high creative quality are expected to be rare and this fits well with positively skewed citation counts as an indicator of creative quality.

The tilted funnel hypothesis: testing the relationship between quantity and quality by means of quantile regression

Moreover, it is important to note that H in the original formulation of the EOB is defined as a subset of T (e.g., the number of high-quality publications are a subset of a scientist's total publications) which implies that the residuals can potentially vary between $-T(1 - \rho)$ and $T(1 - \rho)$. Hence, residual variation is clearly constrained by T and therefore, the EOB implies heteroscedastic errors. In particular, the variance of residuals is expected to increase with T —a pattern that is known in the literature (e.g., Oswald and Johnson 1998). This line of argumentation is illustrated in Fig. 1a. These data were simulated according to the EOB and the scatter around the red regression line clearly follows the pattern of heteroscedasticity implied theoretically by the EOB. In relation to this, from the assumption that H/T is uncorrelated with T , it further follows that the conditional distribution dispersion would scale with the average hit rate ρ . These theoretical predictions suggest that, as compared to the proposed linear relationship between quantity and quality when the conditional mean of quality is predicted, the linear slope to predict the median (i.e., the .50-quantile)

quality can be very similar, or deviate depending on the skewness of the conditional distributions of H . For example, in Fig. 1a; the black dashed regression line to predict the median of the quality outcome is almost identical to the solid red regression line predicting the mean of the quality outcome (for a comparable empirical finding see Michalska-Smith and Allesina 2017). In addition, the linear slopes to predict higher or lower quantiles as compared to the median are expected to differ from the slope at the median (see the regression lines to predict the .10, .30, .70, and .90 quantiles in Fig. 1a). Hence, when extending the EOB from the conditional mean of H to the full conditional distribution of H , a set of specific predictions result that we term the *tilted funnel hypothesis*.

Quantile regression allows for the examination of the linear relation between the predictor variable and the outcome variable at multiple conditional quantiles across the distribution of the outcome variable (Dumas 2018; Koenker 2005; Koenker and Bassett 1978; Wenz 2019; Yu et al. 2003). Quantile regression relies on a weighting procedure and conditional quantile functions that assess the linear relation between the outcome variable and predictor variable at any given quantile of the outcome variable (Koenker and Bassett 1978; Wenz 2019). For example, an ordinary least squares regression between two variables may estimate a significant linear relation, yet a quantile regression assessing the same variables may show the linear relation is non-significant at specific quantiles across the distribution of the outcome variable, or vice versa. Hence, quantile regression provides a flexible and robust approach to study bivariate (and also multivariate) relationships. Indeed, this flexible approach has also been used in the scientometrics literature. For example, Yu and Yu (2016) predicted all deciles of log-transformed average journal impact factor percentile by factors reflecting impact, citable articles, and half-life indices. Shideler and Araújo (2017) predicted the conditional median and the .95-quantile of citation rates of accepted manuscripts as a function of the number of needed reviewer requests. Moreover, the conditional quartiles of time spent on research and other activities were predicted by researcher's age and other demographic characteristics by Kawaguchi et al. (2016).

Heterogeneity in quantile regression slopes is further inherently tied to the presence of heteroscedasticity (e.g., Wilcox and Keselman 2006). In case of homoscedasticity it must be the case that linear regression slopes to predict a conditional quantile q (with q being in the range from 0 to .50) and the conditional quantile $1 - q$ should not be significantly different (Wilcox and Keselman 2006). The EOB, however, implies heteroscedasticity (i.e., residual variation depends on T). Hence, the regression slopes obtained from quantile regression would be expected, based on the equal-odds baseline, to vary when quantity of publications is predicting quality of publications. Moreover, based on the equal odds assumption that H/T and T are uncorrelated, variation of the conditional distribution of H should scale with ρ . Consequently, when we extend the EOB to the full conditional distribution, we would expect a *tilted funnel* shape for the bivariate scatterplot with H on y -axis and T on the x -axis. Larger positive linear coefficients are expected for higher quantiles and smaller positive linear coefficients for lower quantiles (See Fig. 1a for an illustration).

The tilted funnel metaphor is expected to be most apt in datasets that feature average hit rates around 50% (i.e., maximum variance for H/T when ρ ranges between zero and one), and when ρ approaches the hit rate limits (i.e., zero and one), the top or bottom of the funnel will narrow to a line. However, the described funnel shape is only in accordance with the EOB when the full conditional distribution of H/T is independent from T (as it is illustrated in Fig. 1b). In relation to this, it should be noted again that the EOB can be derived based on the assumption that H/T and T are independent (Forthmann et al. 2020a). Hence, the 'tilted funnel hypothesis' implies that we need to test the relationship between H/T and T by means of quantile regression (predicting the conditional distribution of H/T).

and when the found coefficients are rather small (i.e., negligible) or non-existent we would conclude that the data are in accordance with that hypothesis, and the equal-odds baseline theory would be supported.

Aim of the current work

The overarching goal of this work was to thoroughly examine the EOB, (i.e., already known predictions based on the EOB and newly developed ones related to the residuals were under scrutiny). The diversity of scientists and variables in the datasets under investigation provided the opportunity to examine the EOB in multiple scientific contexts. More specifically, we first aimed at establishing the fit of the data to the EOB by examining the following predictions based on previous literature:

- In cadence with Simonton (2003a, b, 2004), non-significant correlations between H/T and T were an indication that the EOB was supported.
- A positive correlation between H and T was expected as a necessary but not sufficient condition for support of the EOB.
- The model-data fit (i.e., the practical usefulness of the EOB) was investigated by means of the SEM approach outlined by Forthmann et al. (2020a).
 - This approach is based on the prediction that, under the EOB, the intercept and slope in Eq. 1 should equal zero and the ratio of the mean of H and the mean of T , respectively (more details are described in the method section below).

In addition, the following novel predictions, developed in this work, were empirically tested:

- In support of the EOB, we expected that the $\Delta_{\hat{\sigma}_a^2}$ would be non-significant in each dataset. In other words, the observed residual variance should not be significantly different from the expected residual variance under the EOB.
- In accordance, with previous findings on the distributions of citation counts, we expected a positively skewed distribution of H/T and a positively skewed conditional distribution $f(H/T)$. Importantly, this expectation is not derived from the EOB. However, as mentioned-above, results related to this expectation help to understand the overall pattern of findings.
- In a quantile regression, the full conditional distribution of H/T should be unaffected by T (i.e., slopes should be near zero at all quantiles).
- Finally, when H is predicted by T by means of quantile regression, it is expected that linear regression slopes would vary, with a positive relationship between the order of the predicted quantile of H and the regression slopes. This prediction further implies heteroscedasticity of residuals resulting from ordinary least squares (OLS) regression with residual variation increasing with T .
 - This heteroscedasticity will also be examined graphically (by means of bivariate scatterplots).
 - This expectation is here termed the *tilted funnel hypothesis*.

Importantly, the tilted funnel hypothesis needs to be understood as an extension of the EOB from a relatively constrained situation in which the conditional mean of H is being predicted (as in OLS regression), to predictions about to the full conditional distribution of

H (as in quantile regression). Analogous to the correlation between H and T , it represents a necessary but not sufficient condition for the EOB. Hence, the *tilted funnel hypothesis* is regarded as subsequent, or downstream from, the EOB, and should not be considered an antecedent of the EOB, and therefore should not be studied in isolation from the nuanced predictions of the EOB.

Method

Data-sources

Dataset 1: inventors

Dataset 1 was taken from the publicly available U.S. National Bureau of Economic Research (NBER) dataset (Hall et al. 2001). More precisely, we used data provided at the Harvard dataverse (<https://library.harvard.edu/services-tolinear/harvard-dataverse>) which were created as part of the Harvard Patent Data Project (Lai et al. 2009; Li et al. 2014) in which each inventor was uniquely identified. This allowed us to analyze patent data at the level of individual inventors. We only used utility² patents from the dataset and excluded data when forward citations were not available. Moreover, potential issues with data truncation were taken into account by using only data from inventors from whom patent data were available exclusively between the years from 1980 to 2003. Hence, inventors with available data prior to 1980 or past 2003 were excluded. Omitting the first five (i.e., 1975–1979) and the last five years (i.e., 2004–2008) for which patent data were available has been identified as a useful strategy to prevent truncation problems with patent data (Dass et al. 2017). This procedure resulted in a final sample of $N=1,520,967$ inventors. The variables analyzed were the number of patents, the overall number of forward citations, and the average citations received over the window of observation (i.e., through 2003). It should be mentioned that we used a subset of the used data here (also the scoring of quality was different) in previous work with a focus on other aspects of the EOB (Forthmann et al. 2019).

Dataset 2: psychological researchers

This dataset was provided by the Leibniz Institute for Psychology Information (ZPID) for the purpose of the current research and was taken from Bauer et al. (2013a). This data has also been used by Bauer et al. (2013b) in a study on the relation between scientific success and both individual as well as organizational characteristics. Dataset 2 included publication and citation data from $N=1742$ psychologists from German speaking countries. The dataset includes 81.6% of the target population of psychology researchers from German speaking countries in the fall of 2010. The variables analyzed from this dataset were the total number of publications and total citations.

² A utility patent is typified by a novel or improved invention with a discernable benefit and use [Inventions Patentable, 35 U.S. Code § 101 (1952)].

Dataset 3: multi-disciplinary scientists

Dataset 3 was obtained³ from Liu et al. (2018), a study that sought to model hot streaks within creative careers (Liu et al. 2018). Specifically, the project examined increases in creative quality during specified time-periods for artists, directors, and scientists. The current study specifically seeks to examine the relation between quality of scientific works and total scientific productions (T) and thus, the dataset containing $N=20,296$ scientists was the only dataset we explored from Liu and colleagues' project. Liu et al. (2018) obtained their sample by disambiguating and merging large-scale datasets obtained from Google Scholar and the Web of Science. Together these databases held roughly 46 million journal articles that were published as far back as 1900 (Liu et al. 2018), making for an incredibly diverse pool of scientists from myriad academic disciplines. This dataset was delimited to scientists having at least a 20-year career and at least 15 publications. The dataset featured a count of total publications for each individual scientist, and the total number citations each individual scientific production received after ten years.

Data analysis

The SEM approach to test the fit of the EOB to a given dataset (Forthmann et al. 2020a) was carried out by means of the *lavaan* package (Rossee 2012) for the statistical software R (R Core Team 2019). The intercept and slope of the simple linear regression of H on T (see Eq. 1) are fixed to zero and the ratio between the sample averages of H and T , respectively. This requires estimation of the mean and covariance structure for H and T (see the OSF for details: <https://osf.io/2zgxn/>). The unweighted least squares estimator with Satorra–Bentler correction (1994) was used, which does not make the assumption of bivariate normality. Fit of the EOB to the data was evaluated by the scaled versions of the RMSEA, SRMR, CFI, TLI, and $\hat{\gamma}$ according to the model fit thresholds provided by West et al. (2012; see the notes on Table 2).

Next, the $\Delta_{\hat{\sigma}_u^2}$ statistic (see Eq. 4) was examined by means of a non-parametric bootstrap approach (Davison and Hinkley 1997) as it is implemented in the R package *boot* (Canty and Ripley 2019). We tested three different estimators for the residual variance: (a) the unbiased variance estimator (dividing by $N - 1$), (b) the maximum likelihood estimator (dividing by N), and (c) the mean sum of squares from the analysis of variance table of the linear regression in Eq. 1. None of these methodological variations changed any of the results and, hence, the pragmatic application of the unbiased variance estimator was chosen. Moreover, we tested the following other statistics to compare the observed residual variance and the expected residual variance: (a) the ratio of observed and expected residual variance, (b) the difference between the log-transformed observed and the log-transformed expected residual variance, (c) the log-transformed ratio of observed and expected residual variance, and (d) the ratio of log-transformed observed and log-transformed expected residual variance. The log-transformation was used with the intention to stabilize the variance of the resulting bootstrap sampling distributions. However, it turned out that the untreated difference as expressed by the $\Delta_{\hat{\sigma}_u^2}$ statistic performed best in terms of symmetry of the sampling distribution across the three datasets (for graphical checks see the OSF repository: <https://osf.io/2zgxn/>). Hence, 95th percentile bootstrap intervals were used for

³ The data is available to the public at: <https://lu-liu.github.io/hotstreaks/>.

Table 1 Descriptive statistics

	Mean	SD
Dataset 1 ($N=1,520,967$)		
Average forward citations	7.60	12.87
Total forward citations	20.81	56.43
Total number of patents	2.47	3.67
Dataset 2 ($N=1742$)		
Average citations	3.71	7.18
Total citations	133.61	262.05
Total publications	35.99	49.35
Dataset 3 ($N=20,296$)		
Average citations after 10 years	26.08	31.30
Total citations after 10 years	1555.76	2365.67
Total publications	55.67	46.20

statistical inference based on $B=1000$ bootstrap samples. Evidence for the EOB would result when the bootstrap interval covers a value of zero.

We further used STATA 16 for regression analyses (StataCorp 2016). In particular, the quantile regression analysis was carried out with the quantile regression functions in STATA (e.g., StataCorp 2013). Parts of the STATA code were inspired by the code made openly available by Sebastian Wenz (2019). The indicators for H , H/T , and T in the respective datasets were all z -standardized prior to regression analyses. Initially, the conditional mean of H (and also of H/T) was predicted by T in an OLS regression. Each analysis only has one predictor, so the OLS regression coefficients (see Table 3 and Table 4) equal correlations and can be interpreted as such. The .10, .30, .50, .70, and .90 quantiles of the conditional distribution of H will be predicted by T in a quantile regression to examine the assumed pattern of heteroscedasticity. In particular, we tested for significant differences between the slopes obtained at each of the five quantiles using the omnibus Wald-test (Koenker 2005). The relationship between H/T and T will also be assessed by means of quantile regression and the omnibus Wald-test with the same set of quantiles as mentioned above. All data preparation and analyses files can be found in the OSF repository (<https://osf.io/2zgx/n/>).

Results

The means and standard deviations for the studied variables, for each of the three datasets included in this study, are shown in Table 1. Next, across the three datasets, correlations between total citations and total publications (r ranged from .59 to .77) were consistently found (see Table 3), which implies that scientists who had more products are likely to be cited more frequently. These findings are visualized by the red regression lines in Fig. 3a through c. These findings are in support of the theoretical position that quality is a linear function of quantity: a necessary but not sufficient condition for the EOB. Moreover, the strength of the correlation between average citations (H/T) and total number of scientific productions (T) was relatively weak (r ranging from $-.0004$ to $.07$) across the three datasets (see Table 4). While the linear regression results for Dataset 2 were non-significant, the linear regression results from Dataset 1 ($\beta=.04$, $p<.001$) and Dataset 3 ($\beta=.07$,

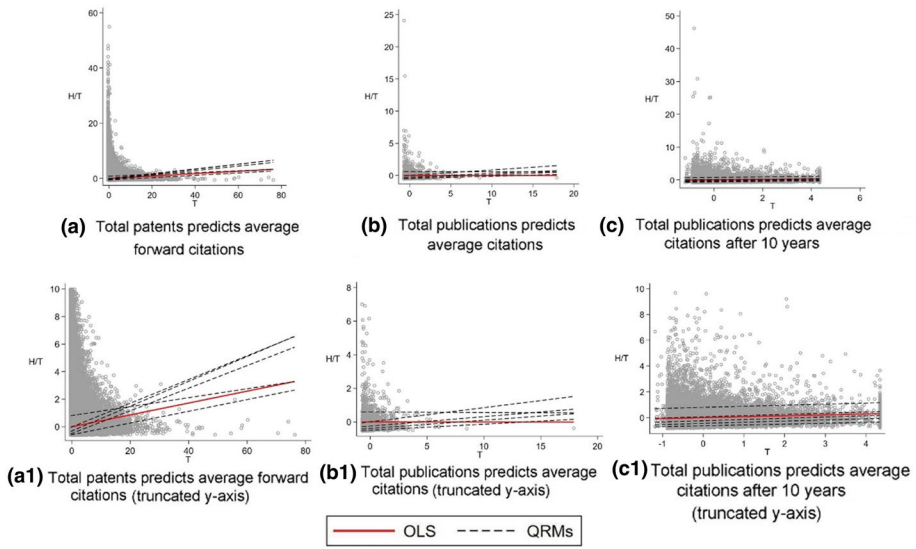


Fig. 3 Bivariate Scatterplots for the relationship between z -standardized H/T (y -axis) and z -standardized T (x -axis). OLS=regression slope from ordinary least squares regression (predicting the conditional mean). QRMs=regression slopes from quantile regression to predict the .10, .30, .50, .70, an .90 quantiles of the conditional distribution. *Top-row* original scatterplots. *Bottom-row* scatterplots with truncated y -axis for better visual inspection of the regression slopes

Table 2 OLS and quantile regression summary: total scientific productions predicts average citations

	OLS coefficients	Standardized quantile coefficients					Omnibus test for quantiles	
		.1	.3	.5	.7	.9	F	p
<i>Predicting the average number of forward citations</i>								
Total number of patents								
Intercept: β_0	.00	-.57***	-.47***	-.31***	-.012***	.82***		
Slope: β_1	.04***	.042***	.082***	.09***	.086***	.032***		
R^2	.0019	.0132	.0178	.0089	.005	.0004	2253.84	< .001
<i>Predicting average number of citations</i>								
Total number of publications								
Intercept: β_0	.00	-.49***	-.38***	-.24***	.009	.59***		
Slope: β_1	-.0004	.036***	.049***	.055***	.084**	-.0032		
R^2	< .00001	.0207	.0099	.007	.0033	< .00001	2.73	.028
<i>Predicting average citations after 10 years</i>								
Total number of publications								
Intercept: β_0	.00	-.63***	-.44***	-.24***	.06***	.80***		
Slope: β_1	.07***	.066***	.09***	.09***	.092***	.083***		
R^2	.0051	.0243	.0206	.0159	.0084	.0023	40.7	< .001

*** $p < .001$, ** $p < .01$

Table 3 Structural equation modeling and residual variance results

	Dataset 1: inventors	Dataset 2: psycho-logical researchers	Dataset 3: multi-disci-plinary scientists
$\Delta_{\hat{\rho}_u^2}$	– 380.10	10,013.98	– 501,254.60
95%-bootstrap interval	[– 429.60, – 335.40]	[– 2925, 24,939]	[– 646,893, – 364,671]
Ratio between observed and expected residual variance	0.830	1.285	0.872
RMSEA	.016	.019	.000
SRMR	.045	.057	.066
CFI	.924	.992	1.000
TLI	.924	.992	1.001
$\hat{\gamma}$.999	.999	1.000

The bootstrap intervals were calculated based on the percentile method and $B=1000$ bootstrap samples. Graphical checks of the sampling distributions of the $\Delta_{\hat{\rho}_u^2}$ statistic are available in the OSF repository (<https://osf.io/2zgxnl>). Cut-off criteria for model fit according to West et al. (2012): RMSEA < .06; SRMR < .08; CFI > .95; TLI > .95; $\hat{\gamma}$ > .95

Table 4 OLS and quantile regression summary: total scientific productions predicts total citations

	OLS coefficients	Standardized quantile coefficients					Omnibus test for quantiles	
		Conditional mean	.1	.3	.5	.7	.9	F
<i>Predicting the total number of forward citations</i>								
Total number of patents								
Intercept: β_0	.00	– .32***	– .24***	– .14***	.02***	.44***		
Slope: β_1	.65***	.10***	.27***	.44***	.69***	1.27***		
R^2	.42	.05	.13	.21	.29	.42	27,824.99	< .001
<i>Predicting total number of citations</i>								
Total number of publications								
Intercept: β_0	.00	– .44***	– .33***	– .16***	.06***	.58***		
Slope: β_1	.59***	.12***	.27***	.49***	.81***	1.49***		
R^2	.34	.08	.17	.24	.32	.44	73.22	< .001
<i>Predicting total citations after 10 years</i>								
Total number of publications								
Intercept: β_0	.00	– .47***	– .31***	– .16***	.06***	.59***		
Slope: β_1	.62***	.21***	.35***	.49***	.68***	1.18***		
R^2	.39	.15	.23	.29	.34	.39	1211.77	< .001

*** $p < .001$

$p < .001$) produced statistically significant associations between average citations and total scientific productions.

In a next step, quantile regression was applied with the total number of scientific productions as the predictor variable and average quality as the outcome variable (see Table 2). Examining the regression slope at each quantile in Dataset 1 and Dataset 3, the strength increased incrementally from quantiles .10 (Dataset 1: $\beta = .0424, p < .001$; Dataset

3: $\beta = .0659$, $p < .001$) to .50 (Dataset 1: $\beta = .0902$, $p < .001$; Dataset 3: $\beta = .0967$, $p < .001$), and then began to wane slightly at quantile .70 (Dataset 1: $\beta = .0861$, $p < .001$; Dataset 3: $\beta = .0917$, $p < .01$) until becoming substantially weaker at or leveling off at quantile .90 (Dataset 1: $\beta = .0324$, $p < .001$; Dataset 3: $\beta = .0826$, $p < .001$). The highest slope for Dataset 1—estimated when the conditional median was predicted—implied that one additional citation on average at the median required ≈ 3 more patents. Figure 3a displays all estimated quantile regression slopes. The highest slope (absolute size) for Dataset 3—estimated when the conditional .70 quantile was predicted—implied that one additional citation at the .70 quantile required ≈ 16 more publications. The omnibus Wald test between quantiles was significant for Dataset 1 and Dataset 3 (see Table 2), suggesting that the regression slope differed from each other. The pattern in Fig. 3a suggests that residuals in Dataset 1 were larger for lower values of T , whereas looking at the scatterplot for Dataset 3 in Fig. 3c the regression lines for each conditional quantile are tightly clustered together (i.e., they appear almost in parallel to the OLS slope).

On average the quantile regression slopes for the prediction of average citations by total publications in Dataset 2 were somewhat smaller as compared to the other datasets. Coefficients increased from quantiles .10 ($\beta = .036$, $p < .001$) to .70 ($\beta = .084$, $p < .01$), and then decreased at quantile .90 ($\beta = -.0032$, $p > .05$). The highest slope (absolute size)—estimated when the conditional .70 quantile was predicted—implied that one additional citation at the .70 quantile required ≈ 76 more publications. Looking at the quantile regression and linear regression lines depicted in Fig. 3b in relation to the distribution of scientists, the differences between quantiles is significant, but only in the slightest (see Table 2), and there is even evidence of quantile crossing between the .70 and .90 quantiles. This indicates that there are scientists in lower quantiles of the total publication distribution that have higher average citations than their peers. All quantile regression slopes were rather small in size and the .9 standardized quantile coefficient was non-significant, meaning that any deviation from the EOB was not strong. The descriptive pattern of residuals in Fig. 3b suggests some larger absolute deviations for points above the OLS regression slope for smaller values of T .

As expected, the distribution of average quality was positively skewed across all datasets (Dataset 1: skewness = 7.55; Dataset 2: skewness = 11.78; Dataset 3: skewness = 12.30). In addition, the conditional distributions of H given T were positively skewed in all datasets. In Dataset 1 we found a positive range of skewness coefficients (1.21–8.47) for inventors with numbers of patents ranging between 1 and 50 (group sizes ranged between $N = 47$ to $N = 882,941$). For the other two datasets a binning procedure to create 20 equally sized groups for quantity was undertaken to examine the skewness of quality in these groups. Again, as expected, positive ranges of skewness were found for both datasets (Dataset 2: 1.27–9.43; Dataset 3: 2.46–21.44).

The $\Delta_{\hat{\sigma}_y^2}$ statistic was significantly different from zero as indicated by the respective 95% bootstrap intervals that did not cover zero for Dataset 1 and Dataset 3 (see Table 3). For Dataset 2, however, the interval covered a value of zero which implies that the observed and expected residual variance under the EOB did not significantly differ from each other (see Table 3). Descriptively, the observed residual variance for Dataset 2 was higher as the expected residual variance by a factor of 1.285 (which equals a reduction by a factor of .778). For Dataset 1 and Dataset 2 the observed residual variance was significantly smaller as compared to the expected residual variance by factors of .830 and .872, respectively (see Table 3). Hence, based on a metric free comparison of observed and expected residual variance, the deviations from the EOB were comparable in descriptive terms across the three datasets. The significant differences found for Dataset 1 and Dataset 2 are likely because of

the very large sample sizes. The latter observation highlights the importance of quantifying the practical usefulness of the EOB to model the given datasets, via model-data-fit.

The SEM fit indices for Dataset 1 were acceptable (CFI and TLI) to excellent (RMSEA, SRMR, and $\hat{\gamma}$) which suggests adequate fit of the EOB to the inventor data (see Table 3). In addition, for the psychological and multi-disciplinary researcher datasets, the fit of the EOB to the data was excellent for all used fit indices (see Table 3). These findings underline the practical usefulness of the EOB as a parsimonious model that fits all of the datasets under investigation here. Among these datasets, the psychological researcher dataset was found to yield the strongest evidence in favor of the EOB (i.e., it is the dataset for which the highest number of expectations of the theory could not be refuted).

Tilted funnel hypothesis results

After establishing the general fit of the EOB to the data, we examined the tilted funnel hypothesis as a necessary but not sufficient condition of the EOB. Significant differences between the slopes across the five quantiles were found for all three datasets (Dataset 1: $F [4, 1,520,965]=27,824.99, p < .001$; Dataset 2: $F [4, 1740]=73.22, p < .001$; Dataset 3: $F [4, 20,294]=1211.77, p < .001$). A closer look at the standardized coefficients exhibited a gradual increase in strength across the whole distribution for all three datasets (Dataset 1: .10 quantile $\beta = .10, p < .001$, .90 quantile $\beta = 1.27, p < .001$; Dataset 2: .10 quantile $\beta = .12, p < .001$, .90 quantile $\beta = 1.49, p < .001$; Dataset 3: .10 quantile $\beta = .21, p < .001$; .90 quantile $\beta = 1.18, p < .001$). These findings suggest a 10% change of quality for a change of quantity by 1SD at the conditional .10 quantile, whereas the same change in quantity would result in a change between 1.2 and 1.5SD at the conditional .90 quantile (i.e., a change of up to 150%). Further, the pseudo R^2 coefficients presented in Table 4 increase gradually over the quantile distribution for all datasets (Dataset 1: .10 quantile $R^2 = .05$, .90 quantile $R^2 = .42$; Dataset 2: .10 quantile $R^2 = .08$, .90 quantile $R^2 = .44$; Dataset 3: .10 quantile $R^2 = .21$, .90 quantile $R^2 = .39$). The strength of the pseudo R^2 coefficients, paired with the significant Wald test findings and evidence that demonstrated the changes in strength of the regression slopes, suggest that the linear relationship between quantity and quality does not hold constant when examining the full conditional distribution. The differences in slope between quantiles were depicted graphically in Fig. 4a–c for each of the datasets, respectively. The quantile regression slopes are represented by the dotted lines, with the bottom line being the .10 quantile and the top being the .90 quantile. These findings are largely in accordance with the tilted funnel hypothesis, associated with the EOB.

Discussion

The EOB is a parsimonious theoretical model for the relation between quantity and quality of scientific productions. In this study, we began with a theoretical investigation, and simulation data demonstration, of the expected residual variance and the ways in which the EOB leads to the empirical phenomenon we describe as the *tilted funnel*. The expected residual variance was examined by means of a bootstrap approach. Then, using three different large datasets related to the creative work of scientists, we used SEM methodology to show that the EOB was in fact a useful and well-fitting theoretical explanation of the covariance patterns in the data. After that, we used quantile

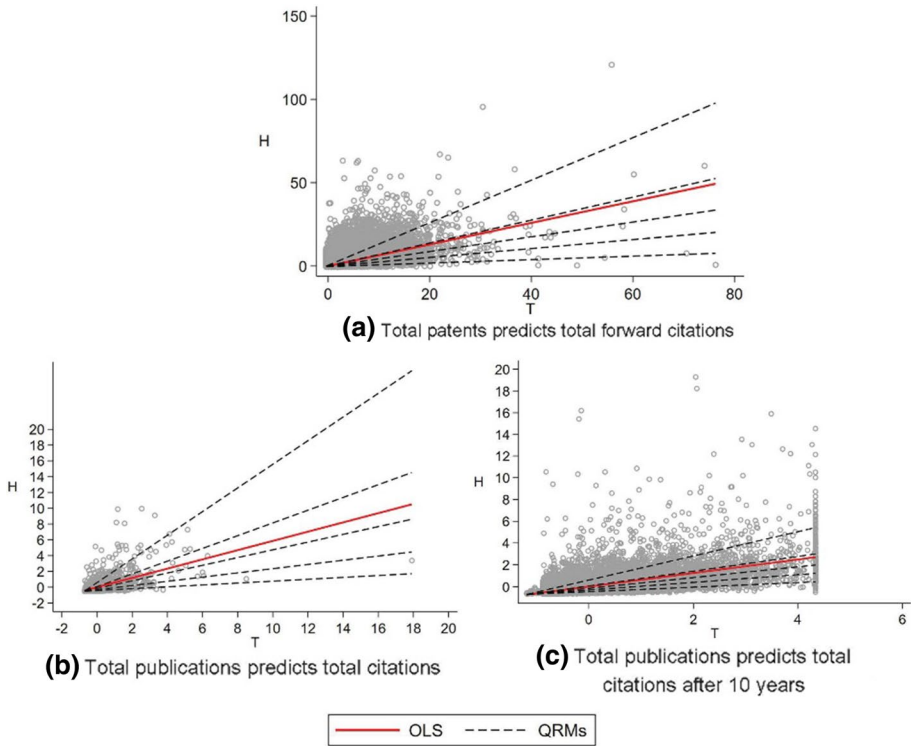


Fig. 4 Bivariate Scatterplots for the relationship between z -standardized H (y -axis) and z -standardized T (x -axis). OLS=regression slope from ordinary least squares regression (predicting the conditional mean). QRMs=regression slopes from quantile regression to predict the .10, .30, .50, .70, an .90 quantiles of the conditional distribution

regression as a promising tool to examine the tilted funnel hypothesis as a theoretical extension of the EOB that takes the heteroscedastic nature of the bivariate relationship between quantity and quality into account.

In particular, we expected that when the total number of citations is predicted by the number of products (i.e., patents or publications), the regression slopes in a quantile regression would increase as a function of the predicted conditional quantile. Heteroscedasticity around the OLS regression line was further predicted to be in accordance with the pattern for quantile regression slopes. In contrast, when the average citations per product were predicted by the number of products, quantile regression slopes were closer to parallel across quantiles, and homoscedastic errors for the OLS regression were assumed. This set of hypotheses was coined the *tilted funnel hypothesis* to provide a metaphor for the expected shape of the bivariate scatterplot based on quantity (x -axis) and quality (y -axis). The tilted funnel hypothesis logically extends important predictions of the EOB theory: (a) the relationship is linear between total products and total citations, (b) the correlation between the average citation count per product and total products is zero, and (c) the residuals should be heteroscedastic in a specific way (i.e., variance of residuals around the OLS slope should increase with the number of products). We further expected and empirically found positively skewed distributions of

H/T and $f(H/T)$ which is well in accordance with the bibliometric literature (e.g., Blagus et al. 2015; Calver and Bradley 2009; Saam and Reiter 1999).

We tested the EOB and the tilted funnel hypothesis using both SEM and quantile regression methods for three datasets for a better generalization of the findings across scientific domains. Dataset 1 and Dataset 2 comprised of inventors generating utility patents and psychological researchers from German-speaking countries, respectively. Dataset 3 included researchers from multiple disciplines and with careers occurring across a large time span. All datasets were rather large which is important for the interpretation of the findings because, in the SEM analysis, sample sizes were substantially large to allow for the generalizability of the covariance patterns from the data to the populations of scientists being studied. The sample size also influenced the quantile regression analysis in that, with this large number of scientists in the data, the full phenomenon of the tilted funnel was able to be observed in the data, with no truncation of the range of the data due to sample size. The residual variance under the EOB was also studied from a theoretical and empirical point of view. It seems that the found deviations of the observed residual variance from the expected residual variance did not exceed an amount that prevented the data to fit the EOB in a SEM approach. Hence, all analyzed datasets displayed findings in accordance with the tilted funnel hypothesis and the EOB. That is, structural models formulated to correspond to the EOB fit the data well or excellently, and among other findings, a positive linear relationship between the number of citations and the number of products was found. In addition, the correlations between average citations per product and the total number of products were estimated to be very small: in the range from $r = -.0004$ to $r = .07$ (importantly, small or practically relevant effect sizes typically range from .10 to .20, especially at large sample sizes; Cohen 1988; Ferguson 2009).

The strongest deviation from this general pattern was found for the multi-disciplinary scientist dataset, where the correlation ($r = .07$) between average citations and total publications implied that 0.5% of the variation in average citations was explained by total publications, and these findings were further corroborated by the found deviation of the observed residual variance from the expected residual variance in the multi-disciplinary dataset. Hence, when taking model parsimony and these weak signs of deviations from the EOB into account, we think that it is fair to conclude that the tilted funnel hypothesis, and the EOB, appear to hold across all datasets and analyses included in this investigation. However, when interpreting the small deviations from the EOB that were found, for example within the multi-disciplinary scientist dataset (i.e., based on the correlation between H/T and T and the $\Delta_{\hat{\sigma}_u^2}$ statistic) or for the inventor dataset (i.e., based on the correlation between H/T and T , the $\Delta_{\hat{\sigma}_u^2}$ statistic, and a slightly deteriorated SEM fit of the EOB), one must take into account that Simonton's theorizing (2010, p. 160) is based on "concepts, facts, theories, laws, hypothesis, formulas, principles, techniques, methods, problems, questions, goals", and so forth that represent the ideas of a research domain (i.e., scientometric researchers contribute to the field of scientometrics, for example). Quasi-random combinations of ideas in a research domain materialize into new scientific productions (Simonton 2009, 2010). Hence, observing the best fit of the EOB to the most homogeneous sample of researchers is well in accordance with Simonton's theorizing. Moreover, one must conclude that the found (yet small) deviations in this study may be in accordance with the dual pathway theory of creativity (that assumes a positive correlation between H/T and T ; see Nijstad et al. 2010): a theory that may warrant further investigation in the future.

Next, it was found that several specific predictions of the tilted funnel hypothesis were supported by the data. First, the quantile regression slopes increased clearly with the order of the predicted conditional quantiles of citation counts across all analyses. Second,

heteroscedasticity was clearly present for all examinations of the EOB by means of OLS regression. Third, the quantile regression slopes when predicting average citations were rather homogenous across quantiles and close to zero in magnitude and, hence, homoscedasticity of residuals for the OLS regressions when predicting average citations was found for almost all analyses (the exception was the inventor dataset).

Limitations

This study is limited in terms of how quality was scored in this study. In line with previous studies that are frequently cited in relation to the EOB (see Simonton 1988, 2003a, 2004, 2009, 2010) we analyzed citation counts as a proxy of creative quality of scientific productions. This view has been supported by other previous researchers (Davis 1987; Wang 2016), but several other conceptions of quality exist that have a theoretical relationship to creativity. For example, there are indicators of originality based on keywords, referenced journals, or cited and citing papers (for an overview of several creative quality measures of scientific productions see Shibayama and Wang 2020). Hence, our study findings may not generalize to other reasonable operationalizations of quality of scientific productions. We argue that, in relation to the EOB, it has been a canonical choice to test the theory by means of citation counts. Thus, extending the presented findings here to other measures of creative quality is an important next step for future work.

It is further important to acknowledge that the EOB is formulated based on counts of high-quality productions (H) and not on citation counts. That is, H is the number of products that are considered to be of high-quality. Davis (1987) makes this distinction by dividing the total publications into those that were cited (H) and those that were not cited ($T-H$), for example. Hence, this can be understood as a weighting of products: high-quality products are weighted by 1 and low-quality products are weighted by zero. Summation of these product weights results in H . For total citation counts, however, the weighting is different because each product is weighted by its number of citations. This methodological issue partly explains differences between the simulated data in Fig. 1a (they are based on counts of high-quality productions) and Fig. 4 that depicts empirical data. The difference is observable mainly for points above the OLS line and rather low values of T . For instances where citation counts are used as the operationalization of H , the residuals are clearly not bounded by $T(1 - \rho)$ and, hence, the pattern of residuals above the OLS line can be different from what would be expected for counts of high-quality products. Residuals below the OLS slope however are still bounded below because citation counts are also restricted to the range of positive integers. In relation to this, we argue that citation counts provide a richer source of information for this study as compared to dichotomization of citations by a median split (see Forthmann et al. 2019), for example. Dichotomization would make the data similar to the original equal odds model formulation, but important information in the data would be lost from the analysis.

A final note relates to varying levels of reliability of average citation scores. That is, the measurement precision of average citations increases with a scientist's total number of products (Cronbach 1941; Dennis 1958; Forthmann et al. 2019) and the analysis used does not take this fact explicitly into account. Forthmann et al. (2019) have used meta-analytical statistical approaches to correct for this source of imprecision in the average citation measures, but in relation to quantile regression we are not aware of a statistical procedure that can handle this issue for citation counts. This lack of reliability was clearly visible in the current study in the regression analyses with average citations as dependent variable.

The large residuals for lower T values above the OLS regression lines in all bivariate scatterplots of Fig. 3 are clearly affected by this issue. Hence, one should take the results for average citations with caution because of potential effects of varying reliabilities of average scores on the slope estimates. Again, this is a very important observation to guide future studies on methodological approaches to investigate the relationship between quantity and quality of scientific productions.

Overall conclusion

The current study provides extended theorizing and empirical tests for the quantity–quality relationship of scientific productions. The assumptions of Simonton's EOB were extended from a focus on the conditional mean of quality to the full conditional distribution of quality. Our extended hypotheses and findings suggest that the linear slope for the prediction of a conditional quantile of the quality (H) distribution increases as a function of quantile order. At the lower quantiles of the distribution of quality, the relationship between quantity and quality is smaller (in this study the linear slope estimates to predict the .10 quantile ranged between .10 and .22) and at the higher quantiles of the quality distribution, the relationship is stronger (linear slope estimates to predict the .90 quantile ranged between 1.18 and 1.60) as compared to the EOB coefficient (linear slope estimates to predict the conditional mean ranged between .59 and .77). We have also highlighted and thoroughly investigated the importance of heteroscedasticity and the expected residual variance in this regard. The existence of other measures of creative quality of scientific productions, the observed differences between counts of high-quality products and citation counts as measures of quality, and the issue with varying levels of reliability of average citations scores include promising possible ventures for future research. Given our findings, we contend that careful attention to the full conditional distribution of scientific quality can and should be a focus in scientometrics research. As this empirical work progresses, we hypothesize that the *tilted funnel* will be apparent in more and more datasets related to the science of scientific production.

Acknowledgements Open Access funding provided by Projekt DEAL. We would like to express our gratitude to Peter Weiland, Armin Günther, and Gabriel Schui from the Leibniz Institute for Psychology Information (ZPID) for allowing us to use the dataset from Bauer et al. (2013a). In addition, we are grateful to Lu Liu for creating the hotstreaks datasets and making them openly available for research purposes. We would further like to thank Carsten Szardenings for writing the simulation function that was used for illustration purposes in this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bauer, H., Schui, G., Krampen, G., & von Eye, A. (2013a). *Besteht ein Zusammenhang zwischen organisationalen und individuellen Merkmalen und wissenschaftlichem Erfolg? Forschungsdaten einer scientometrischen Untersuchung zu psychologischen Forschern in deutschsprachigen Ländern* [How does scientific success relate to individual and organizational characteristics? A scientometric study

- of psychology researchers in the German-speaking countries]. Trier, Germany: Psychologisches Datenarchiv PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation ZPID. <https://doi.org/10.5160/psychdata.brhs11me22>.
- Bauer, H. P., Schui, G., von Eye, A., & Krampen, G. (2013b). How does scientific success relate to individual and organizational characteristics? A scientometric study of psychology researchers in the German-speaking countries. *Scientometrics*, *94*, 523–539. <https://doi.org/10.1007/s11192-012-0760-3>.
- Blagus, R., Leskosek, B. L., & Stare, J. (2015). Comparison of bibliometric measures for assessing relative importance of researchers. *Scientometrics*, *105*(3), 1743–1762. <https://doi.org/10.1007/s11192-015-1622-6>.
- Calver, M. C., & Bradley, J. S. (2009). Should we use the mean citations per paper to summarise a journal's impact or to rank journals in the same field? *Scientometrics*, *81*(3), 611–615. <https://doi.org/10.1007/s11192-008-2229-y>.
- Canty, A., & Ripley, B. (2019). *boot: Bootstrap R (S-Plus) functions*. R package version 1.3-22. <https://CRAN.R-project.org/package=boot>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates Inc.
- Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review*, *32*, 377–390. <https://doi.org/10.2307/2091085>.
- Cronbach, L. J. (1941). The reliability of ratio scores. *Educational and Psychological Measurement*, *1*, 269–277. <https://doi.org/10.1177/001316444100100121>.
- Dass, N., Nanda, V., & Xiao, S. C. (2017). Truncation bias corrections in patent data: Implications for recent research on innovation. *Journal of Corporate Finance*, *44*, 353–374. <https://doi.org/10.1016/j.jcorpfin.2017.03.010>.
- Davis, R. A. (1987). Creativity in neurosurgical publications. *Neurosurgery*, *20*, 652–663. <https://doi.org/10.1227/00006123-198704000-00028>.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.
- Dennis, W. (1958). The age decrement in outstanding scientific contributions: Fact or artifact? *American Psychologist*, *13*, 457–460. <https://doi.org/10.1037/h0048673>.
- Dumas, D. (2018). Relational reasoning and divergent thinking: An examination of the threshold hypothesis with quantile regression. *Contemporary Educational Psychology*, *53*, 1–14. <https://doi.org/10.1016/j.cedpsych.2018.01.003>.
- Feist, G. J. (1997). Quantity, quality, and depth of research as influences on scientific eminence: Is quantity most important? *Creativity Research Journal*, *10*(4), 325–335. https://doi.org/10.1207/s15326934crj1004_4.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532–538. <https://doi.org/10.1037/a0015808>.
- Fischer, J., Ritchie, E. G., & Hanspach, J. (2012). Academia's obsession with quantity. *Trends in Ecology & Evolution*, *27*, 473–474. <https://doi.org/10.1016/j.tree.2012.05.010>.
- Forthmann, B., Szardenings, C., & Dumas, D. (2019). Testing equal odds in creativity research. *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/aca0000294>.
- Forthmann, B., Szardenings, C., & Dumas, D. (2020a). On the conceptual overlap between the fluency contamination effect in divergent thinking scores and the chance view on scientific creativity. *The Journal of Creative Behavior*. <https://doi.org/10.1002/jocb.445>.
- Forthmann, B., Szardenings, C., & Holling, H. (2020b). Understanding the confounding effect of fluency in divergent thinking scores: Revisiting average scores to quantify artifactual correlation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*, 94–112. <https://doi.org/10.1037/aca0000196>.
- Guilford, J. P. (1968). *Intelligence, creativity, and their educational implications*. San Diego: Robert R. Knapp.
- Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological to linear (No. w8498)*. Cambridge: National Bureau of Economic Research.
- Kawaguchi, D., Kondo, A., & Saito, K. (2016). Researchers' career transitions over the life cycle. *Scientometrics*, *109*, 1435–1454. <https://doi.org/10.1007/s11192-016-2131-y>.
- Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.
- Lai, R., D'Amour, A., & Fleming, L. (2009). *The careers and coauthorship networks of US patent-holders since 1975*. Cambridge: Harvard Institute for Quantitative Social Science.
- Lawani, S. M. (1986). Some bibliometric correlates of quality in scientific research. *Scientometrics*, *9*(1–2), 13–25. <https://doi.org/10.1007/BF02016604>.

- Li, G. C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., et al. (2014). Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010). *Research Policy*, 43, 941–955. <https://doi.org/10.1016/j.respol.2014.01.012>.
- Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C., & Wang, D. (2018). Hot streaks in artistic, cultural, and scientific careers. *Nature*, 559, 396–399. <https://doi.org/10.1038/s41586-018-0315-8>.
- Michalska-Smith, M. J., & Allesina, S. (2017). And, not or: quality, quantity in scientific publishing. *PLoS ONE*, 12(6), e0178074. <https://doi.org/10.1371/journal.pone.0178074>.
- Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, 21, 34–77. <https://doi.org/10.1080/10463281003765323>.
- Oswald, F. L., & Johnson, J. W. (1998). On the robustness, bias, and stability of statistics from meta-analysis of correlation coefficients: Some initial Monte Carlo findings. *Journal of Applied Psychology*, 83(2), 164–178. <https://doi.org/10.1037/0021-9010.83.2.164>.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from <https://www.R-project.org/>.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling and more. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Saam, N. J., & Reiter, L. (1999). Lotka's law reconsidered: The evolution of publication and citation distributions in scientific fields. *Scientometrics*, 44(2), 135–155. <https://doi.org/10.1007/BF02457376>.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks: Sage.
- Shibayama, S., & Wang, J. (2020). Measuring originality in science. *Scientometrics*, 122, 409–427. <https://doi.org/10.1007/s11192-019-03263-0>.
- Shideler, G. S., & Araújo, R. J. (2017). Reviewer interest in a manuscript may predict its future citation potential. *Scientometrics*, 113, 1171–1176. <https://doi.org/10.1007/s11192-017-2492-x>.
- Simonton, D. K. (1988). *Scientific genius: A psychology of science*. New York: Cambridge University Press.
- Simonton, D. K. (2003a). Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psychological Bulletin*, 129, 475–494. <https://doi.org/10.1037/0033-2909.129.4.475>.
- Simonton, D. K. (2003b). Genius and g: Intelligence and exceptional achievement. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 229–245). Oxford: Pergamon Press. <https://doi.org/10.1016/B978-008043793-4/50049-0>.
- Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781139165358>.
- Simonton, D. K. (2009). Scientific creativity as a combinatorial process: The chance baseline. In P. Meusburger, J. Funke, & E. Wunder (Eds.), *Milieus of creativity* (pp. 39–51). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-9877-2_4.
- Simonton, D. K. (2010). Creative thought as blind-variation and selective retention: Combinatorial models of exceptional creativity. *Physics of Life Reviews*, 7, 156–179. <https://doi.org/10.1016/j.plrev.2010.02.002>.
- StataCorp. (2013). *Stata 13 Base Reference Manual*. College Station: Stata Press.
- StataCorp. (2016). *Stata Statistical Software: Release 16*. College Station: StataCorp LP.
- Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface*, 4, 707–719. <https://doi.org/10.1098/rsif.2007.0213>.
- Wang, J. (2016). Knowledge creation in collaboration networks: Effects of tie configuration. *Research Policy*, 45, 68–80. <https://doi.org/10.1016/j.respol.2015.09.003>.
- Wenz, S. E. (2019). What quantile regression does and doesn't do: A commentary on Petscher and Logan (2014). *Child Development*, 90(4), 1442–1452. <https://doi.org/10.1111/cdev.13141>.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York: Guilford.
- Wilcox, R. R., & Keselman, H. J. (2006). Detecting heteroscedasticity in a simple regression model via quantile regression slopes. *Journal of Statistical Computation and Simulation*, 76, 705–712. <https://doi.org/10.1080/10629360500107923>.
- Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3), 331–350. <https://doi.org/10.1111/1467-9884.00363>.
- Yu, L., & Yu, H. (2016). Does the average JIF percentile make a difference? *Scientometrics*, 109, 1979–1987. <https://doi.org/10.1007/s11192-016-2156-2>.