Psychonomic Bulletin & Review
2008, 15 (6), 1201-1208
doi:10.3758/PBR.15.6.1201

NOTES AND COMMENT

A hierarchical approach for fitting curves to response time measurements

JEFFREY N. ROUDER
University of Missouri, Columbia, Missouri

FRANCIS TUERLINCKX
University of Leuven, Leuven, Belgium

PAUL SPECKMAN
University of Missouri, Columbia, Missouri

JUN LU
American University, Washington, DC

AND

PABLO GOMEZ
DePaul University, Chicago, Illinois

Understanding how response time (RT) changes with manipulations has been critical in distinguishing among theories in cognition. It is well known that aggregating data distorts functional relationships (e.g., Estes, 1956). Less well appreciated is a second pitfall: Minimizing squared errors (i.e., OLS regression) also distorts estimated functional forms with RT data. We discuss three properties of RT that should be modeled for accurate analysis and, on the basis of these three properties, provide a hierarchical Weibull regression model for regressing RT onto covariates. Hierarchical regression model analysis of lexical decision task data reveals that RT decreases as a power function of word frequency with the scale of RT decreasing 11% for every doubling of word frequency. A detailed discussion of the model and analysis techniques are presented as archived materials and may be downloaded from www.psychonomic.org/archive.

Some of the most influential paradigms use response time (RT) as a dependent measure. In many of these paradigms, the critical finding is how RT depends, quantitatively, on the manipulation. For example, in mental rotation tasks (Shepard & Metzler, 1971), there is a straight line relationship between RT and angle of disparity between the to-be-compared objects. This straight line relationship has motivated the theoretical interpretation of constant velocity mental rotation.

The focus on the functional relationship between RT and manipulations has proved powerful in testing theory even when the relationship is not a straight line. We highlight two examples: skill learning and lexical access. In skill learning, RT decreases as participants gain more practice in a domain. The exact nature of this decrease, whether it follows a power function or an exponential function, or has sharp transitions, discriminates among competing theoretical positions (Anderson, 1982; Dosher & Lu, 2007; Logan, 1988; Rickard, 2004). In lexical access tasks, such as word identification or lexical decision, the time it takes to respond to a word is a decreasing function of word frequency (WF). The form of the relationship between RT and WF also discriminates among competing theoretical positions. Murray and Forster (2004), for example, advocated serial-search models in which the lexicon is ordered by frequency and searched serially. Accordingly, RT is linearly related to the ordinal rank of the word in the lexicon (ordinal rank will be referred to as *rank* throughout). For example, the search time is twice as long for the fourth most frequent word as it is for the second most frequent word. This position can be contrasted to those in which WF effects come about because of changes in processing from repeated exposure to a word. Morton (1969), for example, postulated that repeated exposure to a word lowers the criterial amount of evidence needed for its identification. High-frequency words are encountered more often; hence, responses to them are speeded relative to low-frequency words. Accordingly, RT should depend on WF itself, rather than on rank. In the examples above, WF and amount of practice serve as covariates. In this article, we discuss how to model RT as a function of covariates.

Before discussing our approach to modeling covariates, we review three general properties of RT that seem to hold across many paradigms. We show that if a regression approach does not account for these three properties, the relationship between RT and the covariate may be distorted. Meeting the constraint of these three properties serves as the motivation for the advocated hierarchical Weibull regression model.

Property 1: Participant Variability

It is well known that people vary in all sorts of ways. In fitting curves, it is reasonable to assume that all parameters will vary across individuals. This assumption, however, is violated by aggregating data across individuals. It has been repeatedly demonstrated that aggregating data across people or items may distort the estimate of a functional relationship (Estes, 1956; Haider & Frensch, 2002; Heathcote, Brown, & Mewhort, 2000). Here, we illustrate the problematic effects of aggregation in skill learning. Let T_{ij} denote the ith participant's response to the jth level of practice. Let the true relationship be exponential: $T_{ij} = \theta_{1i} + \theta_{2i} \exp(j \times \theta_{3i}) + \varepsilon_{ij}$, where ε_{ij} is an error term. Figure 1A shows the expected value of $\overline{T}_{\cdot j}$, the aggregated mean, as a function of practice j. The solid line is the case

J. N. Rouder, rouderj@missouri.edu

when there is no variability in rate across participants; that is, for all participants, $\theta_{3i} = \theta_3$. The expected value of $\overline{T}_{\cdot j}$ also follows an exponential with amount of practice j. Figure 1B shows the same relationship in a log-linear plot. In this plot, the exponential relationship corresponds to a straight line. The dotted lines in panels A and B of Figure 1 show the case when θ_3 does vary across individuals. As can be seen, this variability shallows the shape of the curve. This shallowing is especially evident in the upward curvature in Figure 1B. Hence, variability in rate induces the wrong shape in the aggregated learning curves.

Although participant variability is problematic in skill learning, there may be other sources of variability that distort functional relationships. In memory and psycholinguistic inquiries, researchers must also account for variability across items (Baayen, Tweedie, & Schreuder, 2002; Clark, 1973; Rouder & Lu, 2005). The general solution to this problem is to model variability from participants and items as separate from the process under consideration. Hierarchical models are ideal for this purpose, and we introduce an appropriate one for RT subsequently.

Property 2: RT Variance Increases With Means

Researchers typically fit lines and curves by minimizing sum-squared error, and this minimization implicitly assumes that RT has the same variance across the covariate. Figure 2A shows an example of this equal-variance assumption for the power law of word frequency. The three lines show RT densities for WF of 1, 4, and 16. The equal-variance assumption is equivalent to shifting distributions. In Figure 2A, these shifts follow a power function of word frequency. These changes in shift are equivalent to changes in mean; hence, the means decrease as a power function, too. Unfortunately, equal variances are not realistic. Instead, the standard deviation tends to increase with the mean (e.g., Wagenmakers & Brown, 2007). A more realistic model of the same power function on mean RT is depicted in Figure 2B. Here, the effect of WF is to scale the distribution; that is, standard deviation varies as function of word frequency. In panel B, the standard deviation decreases 48% for every doubling of word frequency. Because the standard deviation is linearly related to the mean, the mean also decreases with increasing frequency. In fact, the means follow the same power function as that in panel A.

Does this violation of equal variances matter for deciding between competing functional forms? In order to show the distortions from an inappropriate equal-variance assumption, we ran a small simulation study. Power-function data were generated from the distributions in Figure 2B, which violate the equal-variance assumption. A power function $[E(T_j) = \theta_1 + \theta_2 f_j^{\theta_3}]$ and an exponential function $[E(T_j) = \theta_1 + \theta_2 \exp(\theta_3 f_i)]$ were fit by ordinary least squares. Whereas both models have three parameters, it is reasonable to prefer the one with the lowest sum squared error. We repeated the data generation and minimization 10,000 times. Of these, the exponential had lower squared error 69% of the time, even though the power function was the true model! This artifactual bias toward the exponential comes about because errors from highly variable observations are not downweighted. The bias is such that functional forms appear steeper in shape than they are. The exponential function has a steeper shape than that of a power function. In this case, the steepening bias is sufficient for mistakenly concluding a good fit of the exponential from power-function data. In sum, it is critical to appropriately model error variance in fitting curves to data.

Property 3: RT Distributions Are Shifted

There are two common statistical approaches to relaxing the equal-variance assumption: weighted least squares (WLS) and data transforms. In WLS, observations with high variance are downweighted. In practice, the researcher

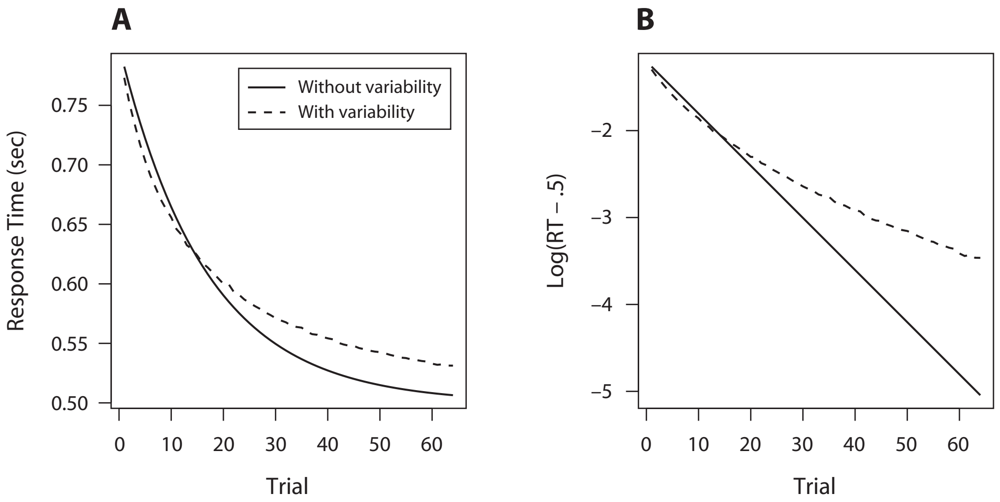

Figure 1. Effect of unaccounted participant variation on rate on the functional relationship between aggregated mean and practice. The true relationship for each individual follows an exponential. (A) Solid and dashed lines denote the cases with and without participant variability in rate, respectively. (B) Log-linear plot showing the true exponential relation as a straight line and the shallowing effect of participant variability as upward curvature.

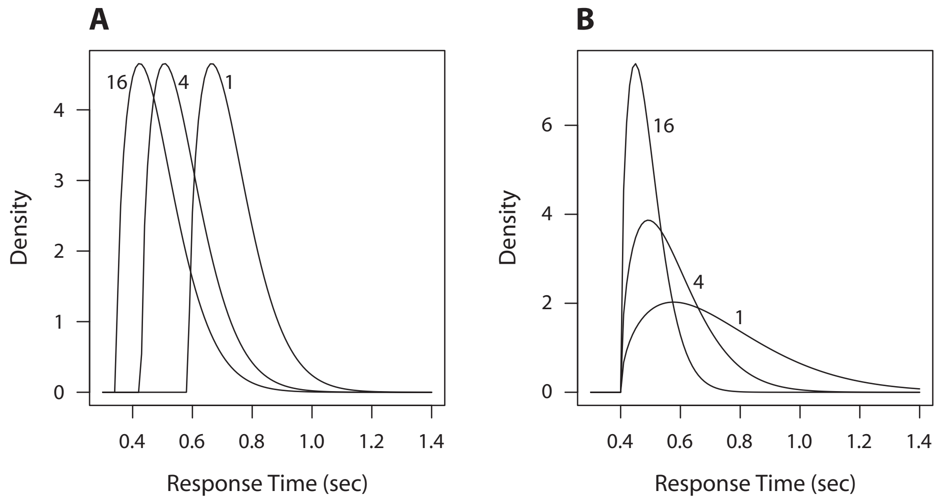

Figure 2. Contrasting effects of word frequency on response time distributions. In both panels, the mean follows the power function $E(T) = .40 + .33f^{-.47}$, where f denotes word frequency. (A) Distributions obey the equal-variance assumption. (B) Distributions for which the mean and standard deviation covary. The effect of word frequency is to decrease standard deviation 48% for every doubling of word frequency.

stipulates a relationship between variance and mean. For RT, a reasonable relationship is that standard deviation increases with the mean. In this case, the predicted means may be used to compute predicted variances, which, in turn, may be used to compute weights. These new weights may be entered into the regression calculation to produce new predicted means. The process continues until convergence and is known as *iteratively reweighted least squares* (Press, Teukolsky, Vetterling, & Flannery, 1992).

A seemingly appropriate data transform for RT is the log function. Social psychologists routinely log-transform RT, especially for significance testing. Of particular interest is the approach of van Breukelen (2005), who recommends placing hierarchical linear models on log RT. Consider, for example, the model

$$\log T_{ij} = \alpha_i + \theta_i c_j + \gamma_j + \varepsilon_{ij}.$$

Parameters α_i and θ_i denote intercept and slope terms, respectively, for the ith participant. Parameter γ_j denotes the effect of the jth item above and beyond that from the covariate. These residual item effects are modeled as zero-centered terms. Parameter ε_{ij} denotes the residual error in log RT. This model meets the two RT properties discussed above: It accounts for multiple sources of variability and stipulates that error variance increases with mean RT. Moreover, it is a fairly conventional hierarchical linear model that is discussed in many texts (e.g., Raudenbush & Bryk, 2002; Skrondal & Rabe-Hesketh, 2004), covered in statistics courses, and conveniently implemented in many packages, such as SAS and HLM. Exponential functions of the covariate result when c_j is the variable of interest (e.g., c_j is the word frequency); power functions result when c_j is logarithmic in the variable of interest (e.g., c_j is log word frequency).

Although both the WLS and the log-transform approach appear appropriate, Rouder (2005) has pointed out a critical problem having to do with the minimum RT. These minima are sizable fractions of the mean. In the quickest of tasks—say, detecting auditory tones—minima and means are about 130 and 180 msec, respectively. In lexical access tasks, the minima and means are about 300 and 700 msec, respectively. These sizable minima require attention, since many process accounts predict minima at 0. The diffusion process in Ratcliff's (1978) diffusion model serves as a suitable example. Regardless of the settings of drift rate or bounds, the minimum of the diffusion process approaches zero as the sample size is increased. To account for sizable nonzero minima in the data, Ratcliff added a constant, T_{ER}, to the latency. This parameter shifts the entire distribution so that the new minimum is T_{ER}. Inclusion of a shift parameter that corresponds to nonzero minimum time is nearly ubiquitous in process-oriented RT models (e.g., Busemeyer & Townsend, 1993; Logan, 1992; Ratcliff, 1978; Rouder, 2000).

Placing linear models on log RT implicitly assumes that RT distributions have minima at zero. Figure 3 shows the effects of the log transform when the data have minimum RT that is substantially different from zero. The data in panel A are generated according to an exponential function with shift. Panel B shows the correct transform, which is to subtract the minimum before taking the log of RT. The fitted curve is a nonparametric smooth (Cleveland, 1981), and its straightness in this plot is diagnostic of exponential relationships. Panel C shows the case of log transforming without subtracting the minimum. Here, the plot reveals upward curvature, which indicates a functional relationship more shallow than the exponential. This shallowing is an artifact of misspecifying the RT distribution as having a minimum at zero.

The need to include nonzero shift parameters in order to account for RT is consequential. From the log-transform perspective, the appropriate transform is log $(T - s)$,

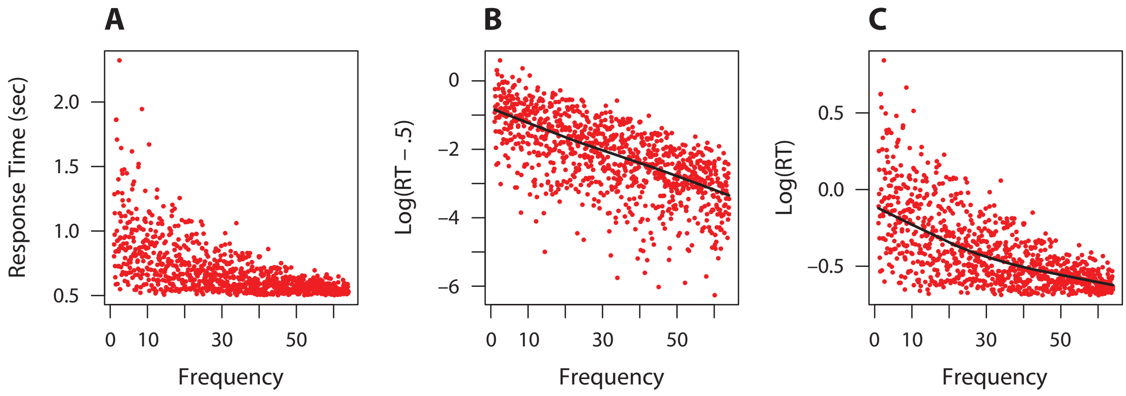

Figure 3. The effect of an unaccounted minimum. (A) Hypothetical observed values from shifted distributions in which the mean obeys an exponential law. (B) Log-linear plot after true minimum has been subtracted correctly shows linearity. (C) Log-linear plots without subtracting the true minimum shows upward curvature indicating a shape more shallow than that of an exponential.

where *s* is a free shift parameter. This transform is outside the GLM family (McCullagh & Nelder, 1989), limiting the appropriateness of common hierarchical modeling packages to regressing RT. Free shift parameters that denote the minimum of a distribution are not accounted for in iterative reweighted least squares regression.

Hierarchical Weibull Models

An appropriate method for regressing covariates on RT should have three properties: (1) It should account for multiple sources of variance, including those from the selection of participants (and items); (2) the standard deviation of RT should grow with the mean; and (3) RT distributions should have free shift parameters that describe the minimum RT. In this section, we present a model that is motivated by these three properties. The model is similar but not identical to our previous Weibull models (Rouder, Lu, Speckman, Sun, & Jiang, 2005; Rouder, Sun, Speckman, Lu, & Zhou, 2003).

Assume that RT on each trial is distributed as a three-parameter Weibull. The Weibull is a flexible distribution with density function

$$f(t\,|\,\psi, \lambda, \beta) = \lambda\beta(t-\psi)^{\beta-1}\exp[-\lambda(t-\psi)^{\beta}],\ t \geq \psi,$$

where ψ, λ, and β serve as shift, rate, and shape parameters, respectively. The roles of these three parameters are shown in Figure 4. The shift parameter in the Weibull corresponds to the minimum RT in the large-sample limit.

The model is extended for many participants and items as follows. Let T_{ij} denote the RT of the ith participant, $i = 1, \ldots, I$ to the jth item,[1] $j = 1, \ldots, J$. The distribution of T_{ij} is

$$T_{ij} \overset{\text{iid}}{\sim} \text{Weibull}(\psi_i, \lambda_{ij}, \beta_i), \tag{1}$$

where the density of the Weibull is given above. In this model, each participant has his or her own shift and shape. The rate parameter is indexed by both participant and item in order to indicate that both of these factors affect rate. This arrangement allows for a great deal of flexibility in modeling variation across individuals while preserving a large degree of parsimony about the effect of item covariates on RT. Rate is an appropriate locus for item effects, because most strength manipulations primarily affect the mean and standard deviation as they do in Figure 4B. Shift is a poor locus for item effects, because changes in shift affect the mean without changing the variance. The rate locus implies that the minimum RT is constant across the covariate. Support for this position comes from Ratcliff and Rouder (1998) and Rouder et al. (2005), who show that shift parameters in the diffusion model and Weibull model, respectively, do not change with stimulus-strength covariates. Further support comes from Andrews and Heathcote (2001), who show that word frequency affects τ in an ex-Gaussian model rather than μ or σ. This pattern implies a relatively constant minimum.

Because rate is a positive multiplicative parameter, it is natural to place a regression model on log rate:

$$\log \lambda_{ij} = \alpha_i + \theta_i c_j + \gamma_j, \tag{2}$$

where c_j represents the value of the covariate for the jth item. Parameters α_i and θ_i denote participant-specific intercept and slope parameters, respectively. Parameter α_i may be interpreted as the speed of the ith participant in naming words with a covariate value of $c_j = 0$; parameter θ_i denotes the multiplier from increasing the covariate. Because participants are randomly sampled, these parameters are modeled as random effects:

$$\alpha_i \overset{\text{iid}}{\sim} \text{Normal}(\alpha_0, \sigma_\alpha^2), \tag{3}$$

$$\theta_i \overset{\text{iid}}{\sim} \text{Normal}(\theta_0, \sigma_\theta^2). \tag{4}$$

The parameter γ_j in Equation 2 denotes all other systematic effects of the jth item that are not accounted for by the covariate. It is assumed to be a zero-centered random effect with variance σ_γ^2. Because participants and items are both modeled as random effects, the model on log λ is an example of a crossed-random effects model (Snijders & Bosker, 1999; van den Noortgate, De Boeck, & Meulders, 2003). The model differs from Rouder et al. (2005), who place an additional log normally distributed error term on λ_{ij}. The previous Rouder et al. (2005) model is more general and computationally more convenient.

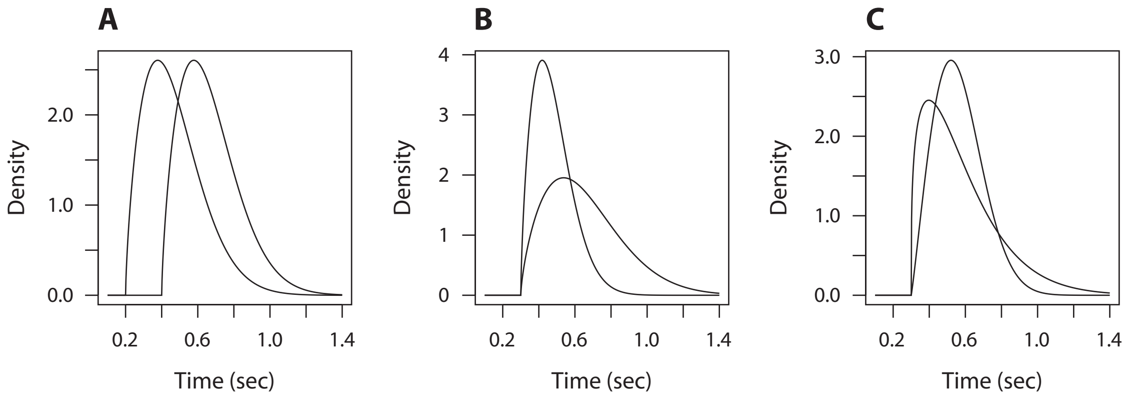

Figure 4. Parameterization of the Weibull distribution. (A) Two Weibull distributions that differ solely in shift (ψ). (B) Two Weibull distributions that differ solely in rate (λ). (C) Two Weibull distributions that differ in shape (β).

Unfortunately, it applies best to cases in which there are replicates of participant–item combinations. In many designs, there are none, and this fact motivated the present model.

Equation 2 provides for a linear relationship between log rate and the covariate. In this case, it can be shown that the mean RT across items for an individual follows an exponential function of c_j:

$$E(T_{ij} \mid \psi_i, \alpha_i, \theta_i, \beta_i) = \psi_i + K_i \exp\left(-\frac{\theta_i c_j}{\beta_i}\right), \quad (5)$$

where K_i is the constant,

$$K_i = \Gamma\left(\frac{\beta_i + 1}{\beta_i}\right)\exp\left(-\frac{\alpha_i}{\beta_i} + \frac{\sigma_\gamma^2}{2\beta_i^2}\right),$$

and where Γ is the gamma function (Abramowitz & Stegun, 1965). The derivation of this fact is provided in the Psychonomic Society Archive of Norms, Stimuli, and Data (www.psychonomic.org/archive).

Alternatively, one might specify a linear relationship between log rate and log c_j. In this case, it can be shown (see the archived materials) that mean RT follows a power function of c_j, given by

$$E(T_{ij} \mid \psi_i, \alpha_i, \theta_i, \beta_i) = \psi_i + K_i c_j^{-\theta_i/\beta_i}, \quad (6)$$

where K_i is given above. This power function has a particularly appealing interpretation: The scale of RT decreases as a constant multiple for every doubling of word frequency.

The three-parameter Weibull model of Equations 1–4 is outside the GLM family (e.g., McCullagh & Nelder, 1989). To make analysis feasible, we adopted a Bayesian framework. Priors on shift and shape are provided in Rouder et al. (2003). Priors on variances σ_γ^2, σ_θ^2, and σ_α^2 are noninformative Jeffreys priors (Jeffreys, 1961). Parameter estimation is performed through Markov chain Monte Carlo integration (Gelfand & Smith, 1990). Further details are provided in the archived materials.

To perform model selection, we used the *deviance information criterion* (DIC; Spiegelhalter, Best, Carlin,

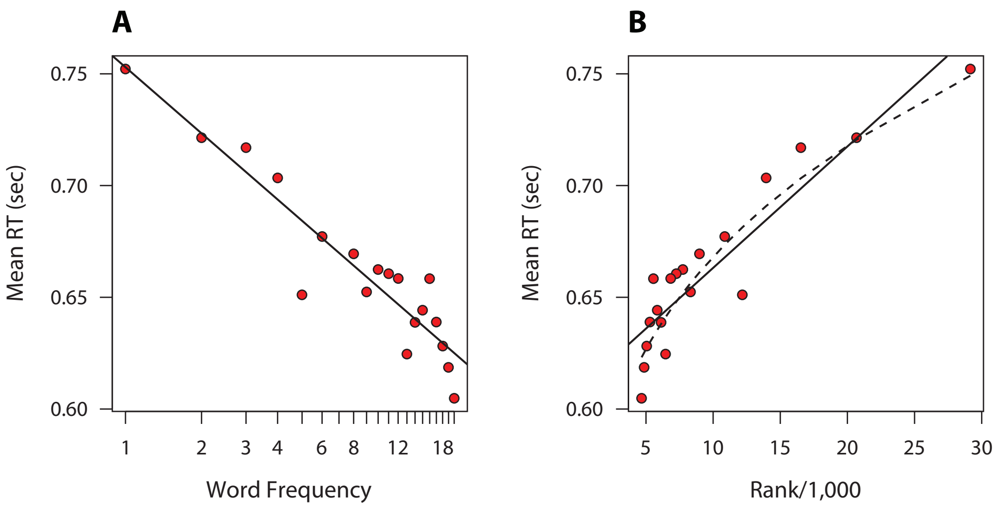

Figure 5. Least-squares regression of mean RT onto logarithm of word frequency (A) and onto rank of word frequency (B). The dashed curved line in panel B shows a power function fit of rank.

& van der Linde, 2002), which is analogous to AIC but designed specifically for Bayesian hierarchical models.[2] Lower DIC values indicate better fit. A secondary means of assessing fit is to plot γ_j as a function of the covariate. Parameter γ_j serves as the item-level residual after regressing the covariate. Because these serve as residuals, it is useful to plot them as a function of the covariate. If a covariate is well specified, these item-residual plots will not vary systematically from zero. If the covariate is misspecified, there will be curvature in the plot.

Application

To illustrate the benefits of hierarchical Weibull regression analysis, we analyzed the lexical decision data of Gomez, Perea, and Ratcliff (2007). Participants performed a go/no-go lexical decision task. If a presented string was an English word, participants depressed a response button; if the presented string was a nonword, they withheld the response. Gomez et al. used 400 word and 400 nonword items. The words varied in Kučera and Francis (1967) frequency from 1 occurrence per million to 20 occurrences per million. A total of 55 participants observed about 180 words each. Figure 5 shows an example of the conventional analysis used by Murray and Forster (2004). Data are mean responses, aggregated across participants and items, as a function of word frequency. Word-frequency covariates are fit to mean RT by ordinary least squares regression. Straight lines fit relatively well for both the logarithm of WF and for rank. The former has served as an empirical benchmark of good fit.

We performed four separate regression analyses with the hierarchical Weibull model: (1) log rate as a linear function of WF (exponential function in WF), (2) log rate as a linear function of log WF (power function in WF), (3) log rate as a linear function of rank (exponential function in rank), and (4) log rate as a linear function of log rank (power function in rank). Figure 6 shows the item residuals, γ_j, for each covariate, as well as the difference in DIC from the best-fitting model. The lines are nonparametric smooths; curvature indicates unaccounted systematic item variability.

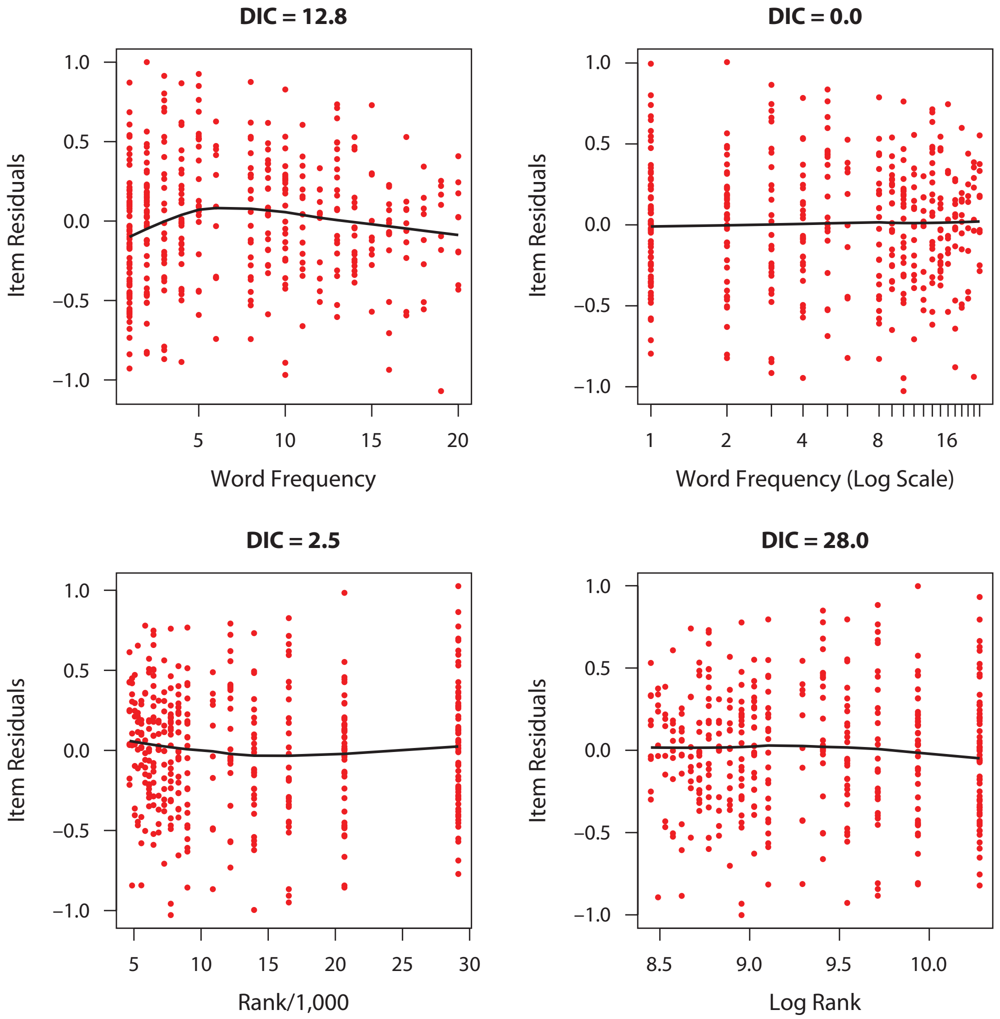

Figure 6. Item residuals (γ_j) for four covariates (WF, log WF, rank, log rank). Lines are nonparametric smooths, and curvature indicates systematic misfit.

The smooths are straightest and the DIC is smallest for the log-WF covariate, indicating that a power function of WF is preferred (albeit mildly). The Murray and Forster (2004) model, which predicts a linear relationship between RT and rank, may be assessed by examining the exponent of the power function resulting from the log-rank covariate. If the linear relationship between RT and rank holds, this exponent must be 1. Across the 55 participants, these exponents range from .23 to .32. Therefore, an exponent value of 1 is not indicated; consequently, the Murray and Forster serial-search model is not preferred. Although we do not recommend using aggregated data to assess competing functional relationships, it is reasonable to ask whether the averaged parameter values account for coarse patterns in aggregated data. The dashed curve in the right panel of Figure 5 shows that a power function with an exponent of .27 does a considerably better job of capturing averaged mean RT data than does rank.

The residual diagnostics in Figure 6 indicate a mild preference for a power function of word frequency. For a participant with average parameters ($\overline{\psi}$, $\overline{\alpha}$, $\overline{\theta}$, $\overline{\beta}$), the mean varies as

$$E(T_j) = .404 + .372 f_j^{-.168}.$$

The value of the average exponent, −.168, corresponds to a decrease of about 11% in the scale of the RT distribution for every doubling of word frequency.

We assessed whether this 11% scaling effect was constant across individuals by examining θ_i, the WF effect on log rate (Equation 2). Figure 7 shows the posterior distributions of these parameters for all 55 participants. The large degree of overlap of these posteriors suggests that this effect is fairly constant; that is, all of Gomez et al.'s (2007) participants showed about the same 11% decline in scale per doubling of WF. The power function has been proposed (McCusker, 1977, as cited in Murray & Forster, 2004), but does not have a theoretical interpretation. The constancy of the 11% reduction per doubling serves as a tantalizing invariance for future theorizing.

The Weibull assumption of the regression model is made solely for convenience. In our experience, the Weibull does not fit RT data as well as competitors such as the log normal or inverse Gaussian do. The problem is that Weibull tails are lighter than the data (Rouder et al., 2005). In the archived materials, we examined the residuals for each observation and found the same misfit that has been previously documented. To ascertain whether this misspecification is important, we ran a number of simulations with data generated from an inverse Gaussian. Even though the Weibull was misspecified, there was good recovery of the underlying functional relationship between the covariate and RT. Hence, the misspecification of the Weibull is not especially problematic for analyzing functional relationships, at least not in this data set.

We caution that the conclusions of the preceding analyses are exploratory. Limitations include the following: (1) The data are from a single task; (2) alternative covariates, such as contextual diversity (Adelman, Brown, & Quesada, 2006) and Zeno word frequencies were not considered; and (3) the range of WF in the Gomez et al. (2007) experiments was small. Further experimentation in combination with hierarchical Weibull regression analyses will provide clarity about the effects of WF on lexical access.

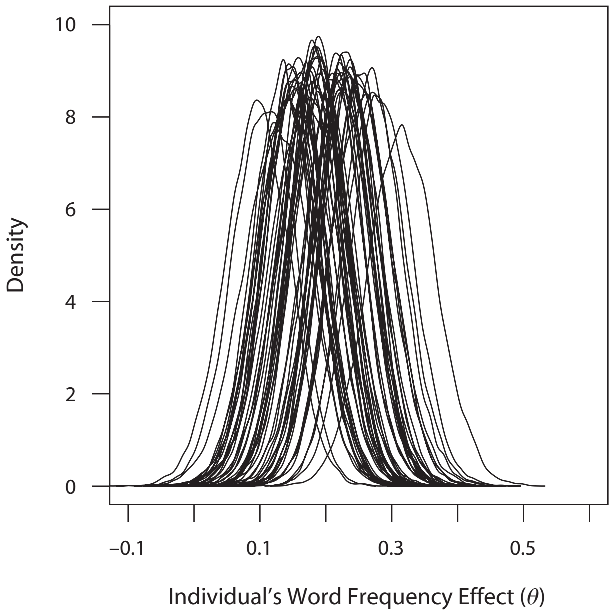

Figure 7. Posterior distributions of θ_i indicate that the word frequency effect is fairly constant across all participants.

Discussion

The hierarchical Weibull regression model is based on three properties of RT: There are multiple sources of variation, RT variance increases with mean, and RTs are shifted. Researchers can embed a number of functional relationships into the regression model by studying how the rate of the Weibull varies across covariates. In the present model, we provide a linear model on log rate of a Weibull distribution. This model is especially well suited for comparing exponential and power functions, which themselves are very general. One functional form that these linear models on log rate cannot capture, however, is the step function that depicts all-or-none learning (Bush & Mosteller, 1951) or transitions from mental states (Rickard, 2004). We are enthusiastic about the Bayesian hierarchical approach taken here because it is likely that this approach will generalize well to a range of nonlinear functional forms, including step functions. In contrast, it is not clear how tractable hierarchical frequentist approaches are for these forms.

AUTHOR NOTE

We thank Richard Morey, Mike Pratte, E.-J. Wagenmakers, and Simon Farrell for helpful comments. This research is supported by Grants SES-0351523 and SES-0095919 from the National Science Foundation, R01-MH071418 from the National Institute of Mental Health, and Fellowship F-04-008 from the University of Leuven, Belgium. Correspondence concerning this article should be addressed to J. N. Rouder, Department of Psychological Sciences, 210 McAlester Hall, University of Missouri, Columbia, MO 65211 (e-mail: rouderj@missouri.edu).

REFERENCES

ABRAMOWITZ, M., & STEGUN, I. A. (1965). *Handbook of mathemati-*

cal functions with formulas, graphs, and mathematical tables. New York: Dover.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, **17**, 814-823.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, **89**, 369-406.

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 514-544.

Baayen, R. H., Tweedie, F. J., & Schreuder, R. (2002). The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon. *Brain & Language*, **81**, 55-65.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, **100**, 432-459.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, **58**, 313-323.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, **12**, 335-359.

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *American Statistician*, **35**, 54.

Dosher, B. A., & Lu, Z.-L. (2007). The functional form of performance improvements in perceptual learning: Learning rates and transfer. *Psychological Science*, **18**, 531-539.

Estes, W. K. (1956). The problem of inference from curves based on grouped data. *Psychological Bulletin*, **53**, 134-140.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

Gomez, P., Perea, M., & Ratcliff, R. (2007). Diffusion model of the go/no-go task. *Journal of Experimental Psychology: General*, **136**, 389-413.

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 392-406.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, **7**, 185-207.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 883-914.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McCusker, L. M. (1977). *Some determinants of word recognition: Frequency.* Paper presented at the 24th Annual Convention of the Southwestern Psychological Association, Fort Worth, TX.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, **76**, 165-178.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, **111**, 721-756.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). Cambridge: Cambridge University Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for decisions between two choices. *Psychological Science*, **9**, 347-356.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 65-82.

Rouder, J. N. (2000). Assessing the roles of change discrimination and luminance integration: Evidence for a hybrid race model of perceptual decision making in luminance discrimination. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 359-378.

Rouder, J. N. (2005). Are unshifted distributional models appropriate for response time? *Psychometrika*, **70**, 377-381.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, **12**, 573-604.

Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, **12**, 195-223.

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, **68**, 587-604.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: CRC Press.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, **64**, 583-639.

van Breukelen, G. J. P. (2005). Psychometric modeling of response speed and accuracy with mixed and conditional regression. *Psychometrika*, **70**, 359-391.

van den Noortgate, W., De Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational & Behavioral Statistics*, **28**, 369-386.

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, **114**, 830-841.

NOTES

1. In skill acquisition, subscript j would denote the level of practice rather than the item.

2. DIC is calculated as $\overline{D} + p$, where $\overline{D}$ is expected deviance and p is the effective number of parameters. Deviance is $D(\Theta) = -2\log[f(y, \Theta)]$, where Θ is the collection of parameters. The expected deviance, $\overline{D}$, is the expectation of D with respect to the posterior distribution of Θ. The effective number of parameters, p, is $\overline{D} - D(\overline{\Theta})$, where $\overline{\Theta}$ is the expected value of the posterior distribution of Θ.

ARCHIVED MATERIALS

The following materials associated with this article may be accessed through the Psychonomic Society Archive of Norms, Stimuli, and Data, www.psychonomic.org/archive.

To access these files, search the archive for this article using the journal name (*Psychonomic Bulletin & Review*), the first author's name (Rouder), and the publication year (2008).

File: Rouder-PBR-2008.pdf.

Description: This file contains a detailed discussion of the model and analysis techniques used in this article, in pdf format.

Author's e-mail address: rouderj@missouri.edu.

(Manuscript received February 5, 2008;
revision accepted for publication June 26, 2008.)