PSYCHOMETRIKA—VOL. 70, NO. 2, 377–381
JUNE 2005
DOI: 10.1007/s11336-005-1297-7

ARE UNSHIFTED DISTRIBUTIONAL MODELS APPROPRIATE FOR RESPONSE TIME?

JEFFREY N. ROUDER

UNIVERSITY OF MISSOURI-COLUMBIA

Van Breukelen (this issue) provides an approach to using both response time (RT) and accuracy for (1) measuring latent abilities of participants even when they may trade speed for accuracy, and for (2) providing insight into the psychological processes underlying task performance. In this commentary, I focus on the second of these aims and assess how useful this approach is for exploring cognition. The approach is based on formulating statistical rather than substantive models on accuracy and RT. I focus here on the appropriateness of the RT component model. This RT model joins a family of recent statistical approaches to RT including those of Glickman, Gray, and Morales (in press); Peruggia, Van Zandt, and Chen (2002); Rouder, Lu, Speckman, Sun, and Jiang (in press); Schnipke and Scrams (1997); and Wenger and Gibson (2004). The advantage of these statistical rather than substantive models is robust estimation and sound inference when data are collected across disparate individuals and items. It is not hard to see how these statistical developments will lead to better substantive theory testing and development in cognitive psychology.

Van Breukelen's model for the ith participant's RT to the jth item is given by:

$$\log(RT_{ij}) \sim \text{Normal}(\mathbf{X}'_{\mathbf{ij}}\gamma_j, \sigma^2), \tag{1}$$

where γ_j is an unknown parameter vector and $\mathbf{X_{ij}}$ is a known covariate vector. Because the model is linear on $\log(RT)$, conventional mixed model techniques may be used to estimate the unknown parameters. Of particular value is that participant effects, item effects, and their interactions may be assessed within the framework.

My colleagues and I advocate that researchers consider attributes of shift, scale, and shape[1] when analyzing RT (Rouder, Sun, Speckman, Lu, and Dzhou, 2003; Rouder et al., in press). Examples of these attributes are given in Figure 1. We argue changes in distributional attributes may imply certain cognitive loci for effects as follows: Changes in shape are paramount and may indicate that the manipulation affects architecture or strategy. If shapes are consistent across a manipulation, then changes in scale may index processing speed. Changes in shift correspond to changes in more peripheral aspects of processing such as encoding stimuli or motor execution of responses. Even if this mapping is incomplete or in error, the disentangling of the distributional loci of effects leads to a natural method of testing cognitive theories. Van Breukelen's lognormal

This research is supported by NSF grants SES - 0095919 and 0351523 to J. Rouder, D. Sun, and P. Speckman. Further support comes from University of Missouri Research Council, Ministry of Education of Spain, and Katholieke Universiteit Leuven. I thank Francis Tuerlinckx for helpful discussions.

Request for reprints should be sent to Jeffrey N. Rouder, 210 McAlester Hall, University of Missouri, Columbia, MO 65211, USA. E-mail: jeff@banta.psyc.missouri.edu.

[1]Concepts of shift, scale and shape can be given precise meaning. Let the density of a random variable exist everywhere and be expressed as $f(t|\Theta_1, \ldots, \Theta_n)$, where $\Theta_1, \ldots, \Theta_n$ are the parameters. Let $z = (t - \Theta_1)/\Theta_2$. We refer to the density as being in location-scale form if there exists some function g such that

$$f(t|\Theta_1, \ldots, \Theta_n) = \Theta_2^{-1} g(z|\Theta_3, \ldots, \Theta_n). \tag{2}$$

If Equation 2 holds, then Θ_1 is referred to as the location parameter and Θ_2 as the scale parameter; parameters Θ_3, ..., Θ_n are shape parameters. When the location parameter indicates the lower bound of support, it is termed the shift parameter.

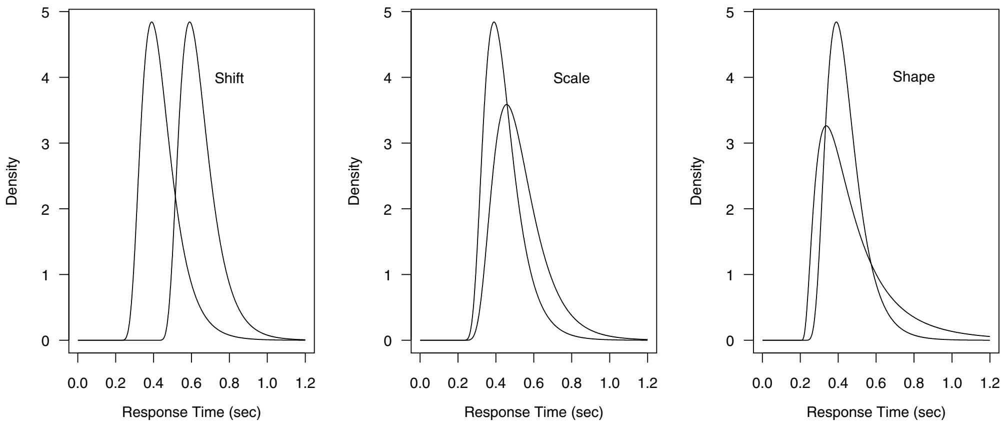

FIGURE 1.
Lognormal densities that vary in shift, scale and shape, respectively.

model fits well within this shift-scale-shape taxonomy.[2] The model is unshifted, or more precisely, has a shift of value zero. There are scale and shape parameters. The effects of covariates, however, are assumed to occur in scale alone. From the vantage of the shift-scale-shape taxonomy, the Van Breukelen model is restrictive. It posits that all effects are on scale. Manipulations can neither change the shape nor shift of a distribution.

How plausible are these restrictions? Consider the empirical evidence about the presence of shifts in RT distributions. Figure 2 shows an analysis from a task in which participants reported the location of a stimulus.[3] The task was trivially easy; performance was nearly perfect and response times were relatively rapid. Eighty observations from each of eighty participants were analyzed. The top panel of Figure 2 shows histograms from two participants who have about the same shape and scale. The difference is in the shift, which is about 0.2 s. This difference is sizeable given that the vast majority of RTs are within 0.6 s of the shift. The bottom-left panel shows the results of fitting an unshifted lognormal (two parameters per participant) vs. fitting one with shift[4] (three parameters per participant). The plot is the empirical cumulative distribution function of participants' log-likelihood test statistics[5] (G^2, Read and Cressie, 1988). If the unshifted model were reasonable, this cumulative distribution function would approximate that of a chi-square with 1 degree of freedom (solid line). It does not. To gain insight into this relatively poor performance, I plotted each individual's G^2 as a function of that individual's estimated shift (bottom-right). The horizontal line indicateds $G^2 = 3.84$, the .05 criterion for the test. Twenty-three of the eighty participants have G^2 higher than criterion; twenty of these participants also have large positive shifts. The poor fit, therefore, is being driven by statistically significant positive shifts.

[2]The two-parameter log-normal in (1) is given by $\theta_1 = 0$, $\theta_2 = \mathbf{X}'_{\mathbf{ij}}\boldsymbol{\gamma}_j$, $\theta_3 = \sigma^2$ and $g(z|\theta_3) = \left(z\sqrt{2\pi\theta_3}\right)^{-1}\exp(-((\log z)^2(2\theta_3)^{-1}))$.

[3]I thank Michael Stadler for use of this data. Observations with RTs greater than 1.2 sec were discarded in the analysis. See Rouder et al. (2003) for methodological details.

[4]Estimation was done by maximizing log-likelihood with the Simplex algorithm (Nelder and Mead, 1965) in the R package.

[5]Strictly speaking, the three-parameter lognormal model is irregular as the bound of lower support is a free parameter. The loglikelihood test may not be distributed asymptotically as a chi-square distribution with a single degree of freedom. A small simulation study revealed that the chi-square distribution with 1 df is an excellent approximation of G^2 for the shapes and scales estimated from the data.

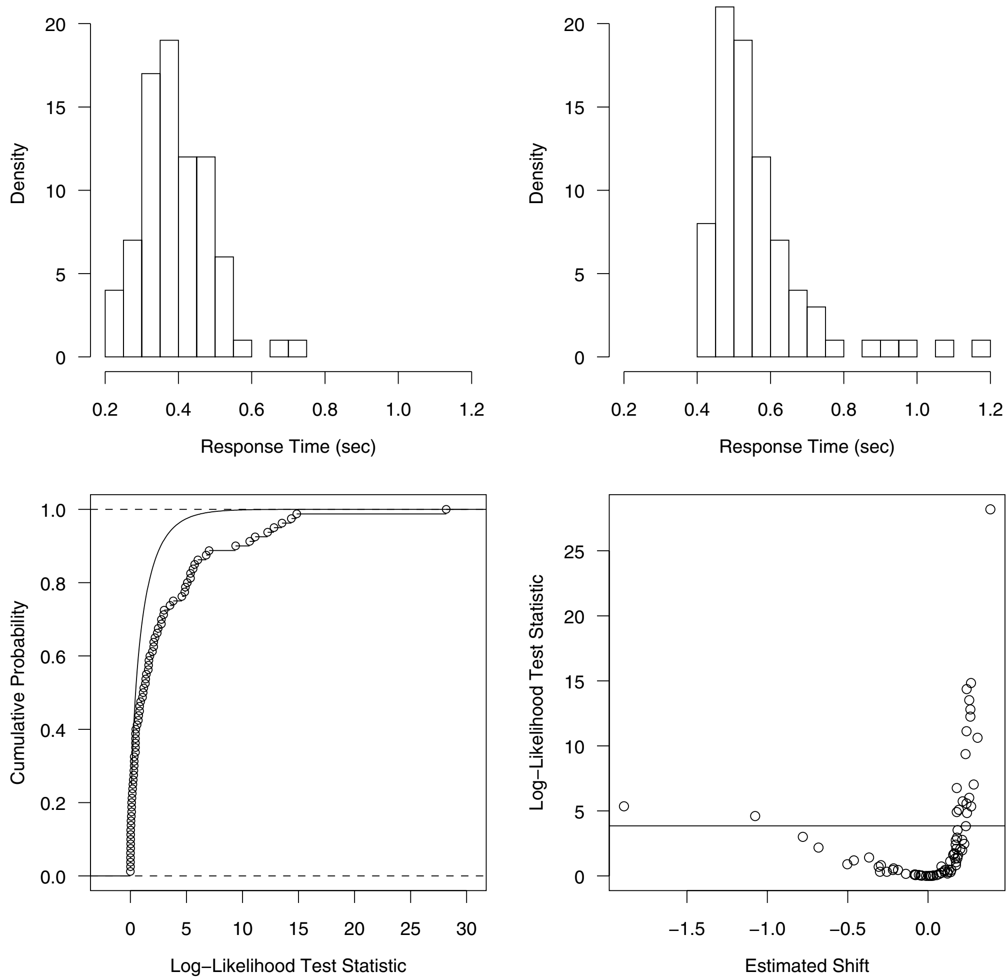

FIGURE 2.
Analysis of shift. *Top panels* show histograms from two participants that vary noticeably in shift. *Bottom-left* panel shows a test of the zero-shift model. Log-likelihood test statistic G^2 is larger than the appropriate chi-square distribution implying a rejection of the model. *Bottom-right* panel shows that the poor performance of the zero-shift model comes about from a preponderance of participants with evidence for shifts greater than zero.

The models in the above analysis are more flexible than Van Breukelen's model—participants were allowed their own shape rather than a single, common shape. As an alternative, I tested the lack of shift for models with common shape. The unshifted, common shape model has a scale parameter for each participant and a single shape parameter; the shifted, common shape model has a scale and shift parameter for each participant and a single shape parameter. The resulting log-likelihood test statistics is $G^2(80) = 377$; which exceeds any reasonable significance criterion. Hence, there is evidence for shifted distributions.

Although I have examined a single data set in detail, the finding of shifted distributions seems to be ubiquitous in the literature. Minimum RT tend to be a sizable fraction of mean RT. For example, minimum RT for auditory detection tends to be about 0.130 s, whereas the mean is about 0.180 s and the standard deviation is about 0.025 s (Luce, 1986). For more complex tasks, e.g., reading words, minimum RT is typically half that of the mean. Many researchers have fit distributional models similar to the shifted lognormal such as the shifted Weibull and the shifted distribution from the first passage time of a diffusion process. In these cases, estimates of the shift range from 0.25 to 0.75 s (see, for example, Logan, 1992; Ratcliff and Rouder, 1998, 2000; Rouder et al., in press).

The cited evidence shows that shifts are nonzero and vary across participants. Overall, shifts tend not to vary with experimental manipulation, but there are notable exceptions. Balota and Spieler (1999), Hockley (1984), and Hohle (1965) document various manipulations that affect only shift. Shifts may also vary systematically across population classes (e.g., Ratcliff, Thapar, and McKoon, 2001). What happens if distributions are shifted and analyzed with an unshifted lognormal? Estimates of scale artifactually increase, and shape (σ^2) artifactually decrease with increasing shift. Such artifacts will certainly influence conclusions about processing.

The presence of shifts greatly complicates the RT lognormal model. Without a shift parameter, the lognormal model can be analyzed with extensions of conventional generalized linear model techniques as documented by Van Breukelen. Unfortunately, these techniques are simply not applicable to shifted models; the shifted lognormal is outside the class of generalized linear models. Van Breukelen's model is not the only one that assumes unshifted distributions. Peruggia et al. (2002) use an unshifted Weibull. Wenger and Gibson's (2004) proportional hazard model implicitly assumes no variability in shifts across participants or conditions. Glickman et al.'s (in press) competing risks model assumes a common shift across covariates. My colleagues and I consider a priori shift restrictions to be tenuous. Consequently, our approach is to allow each item-by-participant combination its own shift (Rouder et al., in press). We find estimation tractable within a hierarchical Bayesian approach (Rouder et al., 2003).

I have not assessed whether the assumption of a common shape (σ^2) is warranted. The evidence about how shape varies across manipulations or participants is not systematically explored in the RT literature. Perhaps the biggest reason for this omission is that it takes thousands of independent and identically distributed trials to estimate shape accurately. My colleagues and I have not assumed constant shape in our Weibull model (Rouder et al., in press). It is unclear, however, whether this flexibility is warranted and a constant shape assumption may be more appropriate in many contexts. We are currently developing Bayes factors for shape parameters to assess this possibility.

In sum, Van Breukelen's general approach of linking statistical models of RT and accuracy will undoubtably be valuable and informative for future substantive development in cognition. I have provided a critique that the two-parameter lognormal is too restrictive to adequately model response times. Instead, shifted versions need be developed, albeit such development is not straightforward as the shifted three-parameter distributional models are outside the generalized linear model family. As shifted models are developed, it will be productive to link them to psychometric models of accuracy.

References

Balota, D.A., & Spieler, D.H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*, 32–55.

Glickman, M.E., Gray, J.R., & Morales, C.J. (in press). Combining speed and accuracy to assess error-free cognitive processes. *Psychometrika*.

Hockley, W.E. (1984). Analysis of reaction time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 598–615.

Hohle, R.H. (1965). Inferred components of reaction time as a function of foreperiod duration. *Journal of Experimental Psychology*, *69*, 382–386.

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.

Luce, R.D. (1986). *Response Times*. New York: Oxford University Press.

Nelder, J.A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.

Peruggia, M., Van Zandt, T., and Chen, M. (2002). Was it a car or a cat I saw? An analysis of response times for word recognition. Case Studies in *Bayesian Statistics*, *6*, 319–334.

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

Ratcliff, R., & Rouder, J.N. (2000). A diffusion model analysis of letter masking. *Journal of Experimental Psychology; Human Perception and Performance*, *26*, 127–140.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effect of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*, 323–341.

Reed, T.R.C., & Cressie, N.A.C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York: Springer-Verlag.

Rouder, J.N., Lu, J., Speckman, P.L., Sun, D., & Jiang, Y. (in press). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin and Review*.

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.

Schnipke, D.L., & Scrams, D.J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*, 213–232.

Wenger, M.J., & Gibson, B.S. (2004). Assing hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 708–719.

Van Breukelen, G.J.P., (2005). Psychometric modeling of response time and accuracy with mixed and conditional regression. *Psychometrika*, *70*, DOI: 10.1007/s11336-003-1078-0.

Manuscript received 17 FEB 2005
Final version received 27 FEB 2005