VALIDATION OF CLASSICAL MAXIMUM LIKELIHOOD ESTIMATION FOR ESTIMATING SHIFTED-WALD MODELS OF RESPONSE TIMES

A Thesis

by

BRYANNA L. SCHEULER

Submitted to the College of Graduate Studies of
Tarleton State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Thomas Faulkenberry, PhD
Committee Members,	Amber Harris Bozer, PhD
	Trina Geye, PhD
Head of Department,	Thomas Faulkenberry, PhD
Acting Graduate Dean,	Nathan Heller, PhD

May 2022

Major Subject: Applied Psychology

ABSTRACT

Scheuler, Bryanna, Validation of Classical Maximum Likelihood Estimation for Estimating Shifted-Wald Models of Response Times, Master of Science. (Applied Psychology), May, 2022, 45 pp., 4 tables, 10 figures, bibliography, 19 titles.

In the psychological sciences, one of the most common ways to measure underlying mental constructs is by studying response times. These response times are fit to a model, and parameters are estimated from that model that provide information about the population. This research project looked at how accurately classical maximum likelihood estimation can determine the parameters from a shifted-Wald model of response times. This was accomplished by performing a parameter recovery from simulated data across five sub-experiments with varying numbers of participants (5, 20, and 80) and trials (20, 80, and 500). Across all five sub-experiments, classical maximum likelihood estimation was largely accurate, with low root-mean square deviations and low mean biases. With the notable exception of when there were only 20 trials, there were moderate to high correlations between the true parameters and those estimated from the model. The number of trials was found to be more impactful on successful estimation than the number of participants. Classical maximum likelihood estimation performed as well with shifted-Wald models as predicted from previous research with other models and could be compared against other parameter estimation methods.

TABLE OF CONTENTS

LIST OF GRAPHICS

TABLE **Page**

FIGURE **Page**

CHAPTER I

INTRODUCTION

The field of psychology takes a scientific approach to measuring and evaluating the human mind and how it functions. But how do we measure a theoretical concept, like happiness? We can't simply use a ruler to measure someone's smile and claim that we are measuring happiness. Instead, we utilize cognitive psychometrics. Cognitive psychometrics is a subfield of psychology that uses models to estimate the latent variables that underly the cognitive processes in question (Wagenmakers et al., 2007). For example, modeling how fast participants respond to different types of math problems can provide information about the steps the brain takes to perform mental arithmetic. Having models that quantify, describe, and predict observed data is what facilitates research in every sector of psychology, bridging the gap between philosophy and science. But how does someone use a model to glean any information about a group of people? Possibly more important, how can we be confident that the model is correct?

Parameter Estimation and Recovery

To answer these questions, it is first essential to understand some basic statistical concepts. Recall that *population* is the entire group of people relevant for the study. For example, an experiment about the sleep habits of college students, the population would be all college students from any university. As it usually infeasible to collect data from every college student, researchers pull a *sample*, or a subset of people selected at random from that population, to represent all college students. To quantitatively represent the observed data in our experiment, we need to convert the observed measurements from our sample into a *statistic*. For the sleep habit example, a statistic might be the average

number of hours slept. These statistics are used to make an inference about what those values would be for a population, which are called *parameters*.

For researchers to understand the collected data from a sample, they need to overlay a model on that data. Metaphorically, this process is like overlapping a Connect-the-Dots puzzle of a dog with a generated image of a dog. While this model may not illustrate every detail of the sample, it does provide a good idea of the overall structure. Generating a model illustrates the relationship between the population parameters and potential observed data. In reverse, beginning with observed data allows us to make a best guess at the parameters, a process known as parameter estimation. Psychologists can use the estimated parameters to address any problems within a population, or even make comparisons to other populations. For example, they may find that college students get proportionally less sleep than working adults. Psychologists could also find trends such as a relationship between a lack of sleep and illnesses, leading them to search for a way for college students to get more sleep. This is just one of countless scenarios where using a model to find a population's parameters can allow psychologists to help people. But how can we know that the method we used for estimating model parameters from observed data works, allowing us to be confident that it gave the *correct* parameter values? The simple answer: with a parameter recovery study.

To explain the concept of a parameter recovery study, consider the following scenario. On a math test, a teacher can generate a problem where they want a specific answer. To evaluate a student's knowledge, the teacher assesses how well the student can find that known answer. This same concept is used when performing a parameter recovery study. A simple framework for parameter recovery was provided by van

Ravenzwaaij and Oberauer (2009), who described four simple steps. First, they generate "target" parameter values from an existing data set. Next, they simulate data based on those parameters. Third, they estimate model parameters from the simulated data. Finally, they evaluate how closely the estimated parameters match the target parameters. By comparing the original parameters to the parameters estimated by the model, a researcher can gauge the accuracy of the method used to estimate the model parameters from observed data. But before we can evaluate if a method of parameter estimation is accurate, we need to address two concepts that help facilitate the 'black magic' of parameter estimation: probability and likelihood.

Probability and Likelihood

One of the key components of a model is its ability to mathematically express the chances of pulling a variety of scores from a sample. Depending on the context and model, these chances can be considered either probability or likelihood. Researchers can make a prediction about these chances, which is called a hypothesis. If a researcher was interested in knowing if there was a relationship between owning a pet and happiness, they could choose to state two possible hypotheses. First, the researcher could say that people who own a pet are not significantly happier than those who do not own a pet. This would be an example of a null hypothesis, H_0, a statement of no relationship between the variables in question. Mathematically, it could be expressed as H_0: $\mu_p = \mu_n$, where μ_p denotes the average happiness for someone who owns a pet, while μ_n denotes the average happiness for someone who does not own a pet. For an alternative hypothesis, the researcher could state that owning a pet changes happiness, be it for the better or for the worse, which could be expressed as H_1: $\mu_p \neq \mu_n$. After stating the hypotheses, the

researcher could collect a sample of happiness scores from pet owners and non-owners and display the differences in happiness in a model. By interpreting the model, the researcher can determine the chance of these scores occurring *if the null is true*, which is called the *p-value*. In other words, the researcher can determine the probability.

A classic example for computing probability is flipping a coin *n* times and landing with the head up *x* times or tails up *n-x* times, where *p* represents the chance of landing with the head up on any trial (Etz, 2018). The probability can be calculated as

$$P(X = x|p) = \left(\frac{n!}{x!\,(n-x)!}\right) p^x (1-p)^{n-x}$$

which calculates the product between the variety of ways to get *x* heads in *n* flips and the probabilities for heads and tails (Etz, 2018, p. 61). For example, we could determine the probability of getting 6 heads in 10 flips when there is an equal probability of either side of the coin:

$$P(X = 6|p = 0.50) = \left(\frac{10!}{6!\,(10-6)!}\right) (0.50)^6 (1-0.50)^{10-6}$$

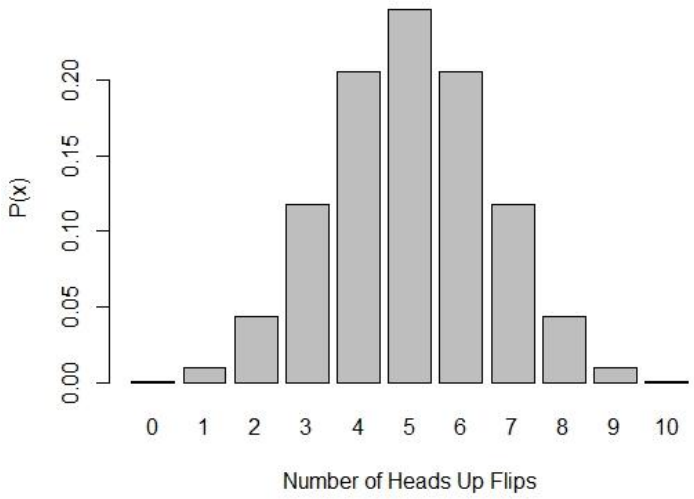
$$= (210) \times (0.50)^6 (0.50)^4 \approx 0.21$$

In this example, all possible outcomes must be between 0 and 10 heads for the 10 flips. In other words, the chances of achieving 0, 1, 2, … or 10 heads must add up to 100% for all possible scores. As such, the probability for *x* scores in *n* flips must sum to 1. In a model, this would be depicted as a graph where the y-axis encompasses all possible numbers of times the quarter could land heads up, as seen in Figure 1. This model can be compared to the actual number of heads obtained in an experiment to determine if the experiment's results are unusual or rare. In terms of the example of pet ownership impacting happiness,

the researcher could use the model to decide if the probability of having that difference in happiness just by chance is so low that they can reject the null hypothesis.

Figure 1

Plot of the Probability of Landing Heads-Up in a Coin Flip



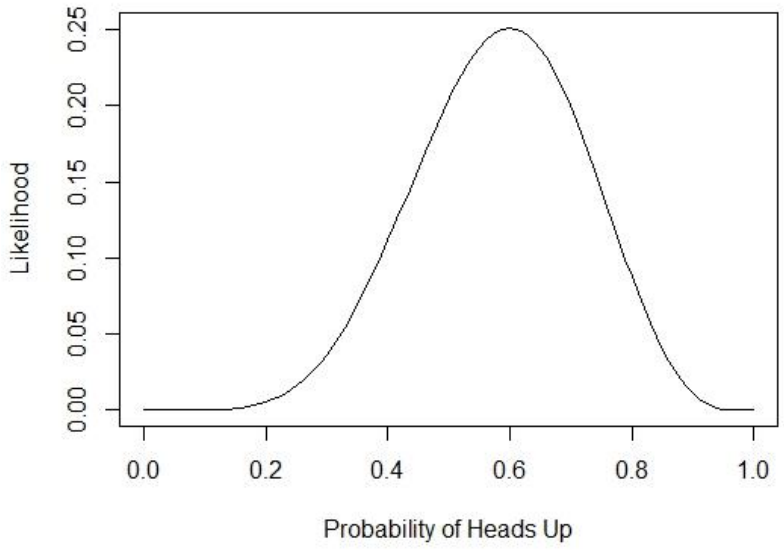
While finding the probability allows a researcher to compare their results to a model when the chance of each outcome is specified, likelihood assesses the plausibility that the observed data were generated by the parameter values in a model. While similar to probability in that it represents the chances of a score, likelihood inverses the approach. Probability is depicted as $P(x|p)$, where there are x number of successes given p chance of success on any one trial, but likelihood can be expressed as $L(p|x)$, or the p chance of success given x number of successes (Myung, 2003). A likelihood example may address, “The coin landed with the heads up six out of the times. How much support

do we have for the chances of heads up being 60%?" In other words, what is the most likely value for the binomial parameter *p*?

When applying this concept, Etz (2018) explained that "the likelihood of a hypothesis *(H)* given some data *(D)* is the probability of obtaining *D* given that *H* is true multiplied by an arbitrary positive constant *K*: $L(H) = K \times P(D|H)$," (p. 60). As the inclusion of this constant makes interpreting one hypothesis alone meaningless, likelihoods are meant to be regarded as a ratio between competing hypotheses. By dividing one hypothesis by another, the constants cancel each other out, and likelihood becomes a ratio of probabilities. This comparative approach can be taken a step further, and instead of looking at two competing hypotheses, a researcher can plot all possible hypotheses simultaneously, as seen in Figure 2. In this graph, "the likelihood ratio of any two hypotheses is simply the ratio of their heights on this curve," (Etz, 2018, p. 62). By this design, we can answer question, "What is the most likely value for the binomial parameter *p?"* This is accomplished by finding the hypothesis with the highest probability simply by finding the highest point in the graph, which is known as the *maximum likelihood*. This maximum likelihood plays a key role in parameter estimation.

Figure 2

Plot of the Likelihood of Heads-Up Probabilities in a Coin Flip



As the delineation between probability and likelihood can be vague in many settings, there are some important differences between the two to clarify from the above given explanations. One of the main differences is that a probability can be interpreted on its own as a blanket statement about the sample, while likelihood must be structured as a comparison between competing hypotheses. Etz (2018) listed a second major difference, as "for conditional probability, the hypothesis is treated as a given, and the data are free to vary. For likelihood, the data are treated as a given, and the hypothesis varies," (p.60). A third difference is that all probabilities must sum to one, while likelihoods have no such constraints due to their nature as a ratio. As all three of these differences have fundamental impacts on the concepts of probability and likelihood, it is necessary to be cognizant when utilizing and interpreting either form.

Classical Maximum Likelihood Estimation

Probability and likelihood are essential components of two of the most common modeling methods: least-squares estimation (LSE) and classical maximum likelihood estimation (CMLE). Myung (2003) explained that LSE, a method that aims to minimize the distance between the data points and the model, may be an inferior method for estimating parameters, as it functions as a descriptive measure rather than providing the inferential capabilities seen in CMLE. CMLE is a method created to find the parameter value with the highest likelihood. Finding the highest likelihood for a parameter is like climbing to the top of a hill—you have reached the point of highest likelihood when moving in any direction decreases your altitude. Similarly, you have reached the maximum likelihood for a parameter's value when changing that value decreases the likelihood, which can be interpreted as finding the true parameter for a population.

CMLE is not infallible, however, and along with the poor applicability in experiments with small trial size designs, one of the main points of concern is the local maximum problem. Myung (2003) discussed how the CMLE algorithm may fixate on a sub-optimal local parameter value, or values, when there is a different global maximum that has a higher likelihood. This would be similar to stating that the top of the hill you climbed to is the highest point, only to later learn that there was a much taller hill a few miles away. While a solution to the local maximum problem has not been created yet, there are methods to avoid getting trapped at a sub-optimal likelihood. Hübner and Pelzer (2020) provided two tactics to circumvent the local maximum problem. For one, a researcher can choose multiple starting points, so that by covering more space, the global maximum is more likely to be found. A second tactic is using previously gained

information to narrow your search to a range of parameter values that are more likely to contain the global maximum, so you don't accidentally get caught on a local maximum. When conducting a CMLE study, it is important to be cognizant of the local maximum problem, so that the parameters you recover reflect the true parameters for the population.

A CMLE estimation can be performed in a software program such as RStudio in three main steps: plotting the data, minimizing the negative log-likelihood, and performing the optimization function (Faulkenberry, 2021). While CMLE is aimed at finding the maximum score, computer optimization algorithms find minimums much easier. Thus, instead of creating a model that looks like a hill, the model looks like a valley. The model has the same x-axis values for each, and the minimum forms a reflection flipped across the x-axis. As multiplying a negative value (like a minimum) by a negative constant turns the value to a positive, we can multiply the minimum scores by negative one to invert the model back to displaying the maximum. As seen previously, computing the chances of a value requires the multiplication of percentages. This presents a problem, as multiplying numbers less than zero has products that get perpetually smaller, to the point that even computers can't compute such small numbers. Thankfully, logarithms can 'save the day' by changing multiplication to addition, allowing the computation of likelihood equations to slowly get larger instead of perpetually shrinking (for more information, see Etz, 2018; to learn how logarithms can be used with categorical data, see Agresti, 1989). These concepts are combined in a process known as finding the negative log-likelihood. After entering the data, we find the minimum likelihood by oscillating around the parameter values that are least likely, shortening the intervals until it arrives at the minimum, similar to a playground swing

that slows to a stop at the lowest point. We can then flip the model by taking the negative log-likelihood. In the final step, we create an optimization function that includes the data, the negative log-likelihood function, and an initial guess for what our parameters should be (which can help avoid the local maximum problem). Running this optimization function finds the parameters that maximize the likelihood function for the given data set. The model created with these parameters can be plotted on a data set to illustrate how well the model fits the data.

Types of Models

In psychology, one of the most common variables to measure is how long it takes participants to take in information and respond, which is called a response time (RT). In a model, these RTs tend to be unimodal (where the data rises towards one peak, then decreases again) and positively skewed, where the data clusters around the faster RTs, then slowly tapers off as the time to respond increases. Matzke and Wagenmakers (2009) stated that the field of cognitive psychology is beginning to see the importance in the entire shape of the RT distribution, rather than only being concerned with the average. This is largely due to the underlying cognitive processes that impact how long it takes a person to respond. Schwarz (2001) explained, "the decision to initiate and execute a specific overt response does not arise holistically, in an all-or-none fashion, but is rather preceded by a stage during which response-related information gradually accumulates over time," (p. 459). Simply stated, the human mind drifts to a decision, and evaluating response times in a sufficient model can provide extensive information about the cognitive processes involved.

One of the models used to depict response times is the ex-Gaussian model. This distribution combines a bell curve with an exponential curve that forms a tail, creating a shape that fits very well with response times. The parameters for these models are the mean, μ, standard deviation, σ, and the exponential tail component, τ. The area under the curve, or the probability density, is found by the function

$$f(x|\mu,\sigma,\tau) = \frac{1}{\tau\sqrt{2\pi}} exp\left(\frac{\sigma^2}{2\tau^2} - \frac{x-\mu}{\tau}\right) * \int_{-\infty}^{[(x-\mu)/\sigma]-(\sigma/\tau)} exp\left(-\frac{y^2}{2}\right) dy$$

as explained by Matzke and Wagenmakers (2009). While this model fits RTs very well, the model does have drawbacks. The most impactful of these is the lack of plausible theoretical rationale in interpreting its parameters. The cognitive processes that are attributed to μ, σ, and τ vary vastly between different researchers (for examples, see Matzke & Wagenmaker, 2009, p.800). As such, the ex-Gaussian model can be used for descriptive purposes but is not ideal for making inferences about cognitive processes.

Though descriptive models may not be well attuned to making inferences, not all types of models are similarly constrained. This second type of model, sometimes called accumulator models, are noteworthy in their ability to describe an entire distribution of parameter values as well as provide information about the underlying cognitive processes (Faulkenberry, 2017). While there are many types of accumulator models, including the drift diffusion model (Ratcliff & Murdock, 1976; Alexandrowicz & Gula, 2020) and the EZ-diffusion model (Wagenmakers et al., 2007), the accumulator model that this study focused on was the shifted-Wald model. Anders et al. (2016) remarked that the shifted-Wald can provide descriptive information about the parameters for a distribution while providing "a simple activation accumulation model for the observed responses in a given paradigm," (p. 310). While the shifted-Wald can be generalized for any experimental

domain with positively valued data with a slight right skew, the most common use for this model is RTs, making it a good candidate for the present parameter recovery study.

The shifted-Wald model is composed of three parameters: shift (θ), drift rate *(γ)*, and response threshold (α). Shift refers to the pocket of time allotted for observing and processing the stimuli, before beginning the response process. Drift rate is the rate of accumulation of stimulus information, and typically reflects test difficulty. Response threshold is the amount of information needed to make a response. These parameters are reflected mathematically in the following probability density function

$$f(x|\alpha,\theta,\gamma) = \frac{\alpha}{\sqrt{2\pi(x-\theta)^3}} * exp\left\{-\frac{[\alpha-\gamma(x-\theta)]^2}{2(x-\theta)}\right\}$$

as explained by Matzke and Wagenmakers (2009). While the parameters' associations with the underlying cognitive processes sets shifted-Wald models apart from ex-Gaussian models, there are still limitations seen in shifted-Wald models. A variety of studies have found that shifted- Wald models may be better suited for situations with a single response boundary, such as simple go/ no-go tasks, as it is still unclear how well shifted-Wald models can be applied to more complex designs (Matzke & Wagenmakers, 2009; Miller et al., 2018). Due to these constraints, this study was designed with a single response boundary. Given the potential importance of the parameter values in a shifted-Wald distribution for modeling response times in cognitive psychology tasks, it is essential that we are able to trust the parameter values we obtain when using classical maximum likelihood estimation. To validate CMLE as a fitting method, I performed a parameter recovery study.

The Present Study

In a heavily cited study, Farrell and Ludwig (2008) performed a parameter recovery study evaluating four approaches to modeling response times: classical maximum likelihood (the CMLE method mentioned previously), single-level Bayesian (SLB), hierarchical maximum likelihood (HML), and hierarchical Bayesian (HB). By applying these approaches to a series of experimental designs with simulated data, Farrell and Ludwig (2008) were able to compare the accuracy of each approach for modeling response times in an ex-Gaussian model for different sample sizes and number of trials. To perform an experiment with simulated data, the researchers used parameters from parent ex-Gaussian distributions that approximated the means used by Cousineau et al. (2004). They used these parameters to generate random parameters (mean, standard deviation, and a tail component that reflects information processing) for participants under the assumption that individuals would be randomly distributed throughout the population. Farrell and Ludwig (2008) created five sub-experiments to illustrate combinations of varying sample sizes (5, 20, or 80 participants) with different numbers of trials (20, 80, or 500 trials). After simulating data sets for the different sub-experiments, Farrell and Ludwig (2008) analyzed the four modeling methods by their accuracy, mean bias, and correlation, and included scatterplots to visualize. They found that the two hierarchical models tended to be more accurate, especially amongst smaller sample sizes. Similarly, the hierarchical models tended to show less mean bias than the single-level methods, with less-variable estimates. Interestingly, all four methods had strong correlations for mean and standard deviation, with only slight differences for the tail component. Between the two single-level approaches, SLB was more accurate for smaller amounts of trials, while CMLE was more accurate for large amounts of trials. While SLB

was more biased than CMLE, it had less variance. Across the correlations for all three variables, CMLE tended to have the weakest correlation, but rarely by any vast difference. Altogether, CMLE tended to be on the lower end of performance amongst the four methods. But does this mean that CMLE is an unsatisfactory method for parameter recovery?

The purpose of this study was to evaluate CMLE as a method for parameter estimation with a shifted-Wald model of response times. I utilized the same approach as Farrell and Ludwig (2008) by simulating data from a parent population, then testing the method on the same variety of sample sizes and numbers of trials. After simulating the data, I evaluated the accuracy, mean bias, and correlations in the parameters (shift, drift rate, and response threshold) between the estimated parameter values and the "true" target values I generated the simulated data from. I hypothesized that CMLE would accurately model the response times, with low mean biases, and strong correlations, providing support for it being a valid estimation method.

CHAPTER II

METHOD

Simulation Design

I simulated data in a free software program called RStudio so I could compare the true parameter values to those estimated by CMLE. Just as a teacher needs to know the correct answer to grade a student, to evaluate how well CMLE estimates a parameter, a researcher needs to know the true parameter. To avoid the local maximum problem, I chose to simulated data from variables that were representative of typical shifted-Wald parameter values. These parameters came from a study of the size-congruity effect by Faulkenberry et al. (2018) that utilized shifted-Wald models to plot response times. The values I selected came from one category of trials, where Faulkenberry et al. (2018) reported a mean drift rate $\gamma = 3.91$ ($SD = 0.70$), a mean response threshold $\alpha = 0.92$ ($SD = 0.17$), and a mean shift $\Theta = 0.32$ ($SD = 0.05$).

The first step in simulating data for this experiment could be thought of as creating "artificial people." The parameter values from Faulkenberry et al. (2018) were averages from a sample, but people are unique, so their response times (and corresponding shifted-Wald values) will vary around the mean. In fact, the artificial people should have shifted-Wald values normally distributed around the averages reported by Faulkenberry et al. (2018) for each parameter value. Thus, I started the simulation process by creating normal curves for drift rate, response threshold, and shift. Each artificial participant had a value randomly drawn from each curve. These three values were the true parameter values for each participant.

Just as response times can be broken into the three shifted-Wald components of drift rate, response threshold, and shift, combining the three components can create the response times for each subject. Thus, for the second step in simulating data, I used each artificial participant's values to create their own distribution of response times. By generating these response times, my data was in the same format that a 'normal' experiment would have collected with real participants.

With distributions of response times for my artificial participants, I could apply CMLE just like any other experiment. Therefore, my third step of the simulation process fit shifted-Wald models to my distributions of response times, and estimated the drift rate, response threshold, and shift parameters for each participant. I then compared these estimated parameters to the true parameters I created in the first step of simulating data. This comparison was made for every participant in a sample, and the averages for each sample were recorded. I then repeated this process 1,000 times, divided evenly between five sub-experiments with different experimental designs. While I adjusted both the number of participants in each sample and the number of response times, or trials, "collected" from each participant for different sub-experiments, the process of simulating data remained the same throughout the entire process.

In replication of the setup used by Farrell and Ludwig (2008), I simulated data in five sub-experiments with 200 samples each. In the first experiment, there was 20 participants with 20 trials per participant. In the second experiment, I changed the number of trials to 80. In the third experiment, the number of trials was increased to 500. These three experiments were designed to illustrate how well the CMLE estimates a shifted-Wald model for small, medium, and large trial sizes, respectively. To test CMLE

in studies with very few participants but lots of trials, the fourth experiment had 5 participants with 500 trials each. The final experiment tested CMLE in studies with lots of participants, but fewer trials, by simulating 80 participants with 20 trials per participant. These five sub-experiments evaluated CMLE's validity in the variety of situations where CMLE may be used.

Evaluation Methods

I evaluated how well CMLE recovered parameters from a shifted-Wald model utilizing both metrics and visualizations. To quantify the adequacy of CMLE, I calculated the root-mean square deviations, mean bias, and the correlation between the true parameters and the estimated parameters. For visualizations, I created correlation scatterplots and box plots of the means for my results.

To first visualize the results, I created boxplots for each of the three parameters. These plots illustrate the mean, minimum, and maximum for each parameter. In addition, the statistics program found the points directly between the minimum and mean, then between the mean and the maximum. These two points create the interquartile range, which represents the middle 50% of the data. When comparing these boxplots to the true parameters, it is easier to see how well the model encapsulates the true parameter values. The goal was for the true parameter values to fall in the middle of the boxplots.

For the first metric, I calculated the root-mean square deviations (RMSD) for the simulated data. The RMSD measured the differences between the true parameters and those generated in the model with CMLE by calculating

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}(x_t - x_e)^2}{N}}$$

where x_t = true parameter values, x_e = CMLE estimated parameter values, and N = number of values. The goal was for the model to fit the data as closely as possible. Thus, the more accurately CMLE recovered parameter values, the lower the RMSD. This evaluation method was applied to the drift rate, threshold, and shift for all five sub-experiments, providing a way to compare the accuracy of CMLE amongst different sample designs to provide a more robust understanding of CMLE's applicability. When utilizing this approach, Farrell and Ludwig (2008) found higher RMSD values that ranged from 0.0297 to 0.0478 in sub-experiments with 20 trials per participant, but lower RMSD values when there were 500 trials, ranging between 0.0045 and 0.0075. While this study applied CMLE to shifted-Wald distributions rather than ex-Gaussian distributions, I predicted that this trend in RMSD values for different sample designs would still be observed.

Next, I estimated the mean bias of the CMLE estimated parameter values. The error of the model that can be attributed to mean bias in the model is found by

$$MB = \frac{1}{N}\sum_{i=1}^{N}(x_t - x_e)$$

where x_t = true parameter values, x_e = CMLE estimated parameter values, and N = number of values. If the value for mean bias was positive, the model had overestimated the parameter values. Alternatively, a negative mean bias illustrated that the model underestimated the parameter values. Thus, a more favorable mean bias will be as close to zero as possible. Just as was done for RMSD, mean bias was found for all three parameters for each of the five sub-experiments to demonstrate CMLE's ability to create a model for different sample designs. Furthermore, the bidirectional nature of mean biases illustrated if CMLE systematically overestimated for one variable or

underestimated for another. Farrell and Ludwig (2008) found mean biases as large as large as 0.0127 in samples with only 20 trials and as small as 0.0001 in samples with 500 trials, a trend I expected to hold true for shifted-Wald models.

For the final metric evaluation of CMLE in shifted-Wald models, I found the correlations between the true parameter values and those generated by CMLE. Correlation is the degree to which two variables are related. There are a variety of types of correlation based on the constraints and variables of the data. This study evaluated response times, a variable that begins at zero and continues indefinitely, which is called a continuous variable. As such, a Pearson correlation was utilized for this data. The Pearson correlation was calculated as

$$r = \frac{\sum(x_t - \overline{x_t})(x_e - \overline{x_e})}{\sqrt{\sum(x_t - \overline{x_t})^2 \sum(x_e - \overline{x_e})^2}}$$

where r = the coefficient for the correlation, x_t and x_e = the values from the true and CMLE estimated parameters, $\bar{x}_t$ and $\bar{x}_e$ are the averages from their respective parameters. A positive correlation coefficient illustrates that both variables increase simultaneously, whereas a negative correlation represents that as one variable increases, the other decreases. The closer the correlation coefficient is to positive or negative one, the stronger the relationship, whereas a correlation close to zero represents a weak relationship. In their shifted-Wald study with 30 participants who had 120 trials, Steingroever et al. (2021) found strong, positive correlations for all three parameters, but they noted that drift did have notable variance. I predicted that the CMLE parameters would highly correlate with the true parameters for this study as well.

To aid in visualization, these correlations were created in a scatterplot, illustrating the relationship between the true parameters and those obtained by CMLE. A line of best

fit was applied to the data, so the viewer could see how far each dot is from the ideal. As Steingroever et al. (2021) explained, “under perfect parameter recovery, all dots would lie on the main diagonal … indicating little uncertainty about the recovered parameter values,” (p. 1,064). This design allowed the reader to understand how close the model was to the ideal fit.

CHAPTER III

RESULTS

Once the data had been simulated, RStudio was utilized to create visuals of the data, and find the RMSD values, mean biases, and correlations between the true parameters and those estimated by CMLE for the 200 samples in each of the five sub-experiments. The CMLE estimated means are depicted in Table 1. When the number of trials was at least 80, the averages of the CMLE estimates were very similar to the original values from Faulkenberry et al. (2018), which serve as the averages for the true parameters. In fact, for both experiments with 500 trials per subject, the averages of the CMLE estimated parameters were nearly identical to the true parameter averages. Notably, despite quadrupling the number of participants in the fifth sub-experiment, the averages of the CMLE estimates were no closer to the true parameter averages than the first sub-experiment.

Table 1

The CMLE Estimated Means and Standard Deviations (SD)

Experiment	Gamma (γ)		Alpha (α)		Theta (θ)	
	Mean	SD	Mean	SD	Mean	SD
P = 20, N = 20	4.49	0.60	1.88	0.98	0.27	0.06
P = 20, N = 80	3.97	0.23	0.95	0.08	0.32	0.01
P = 20, N = 500	3.91	0.17	0.92	0.04	0.32	0.01
P = 5, N = 500	3.92	0.32	0.92	0.09	0.32	0.02
P = 80, N = 20	4.52	0.31	1.89	0.45	0.27	0.03

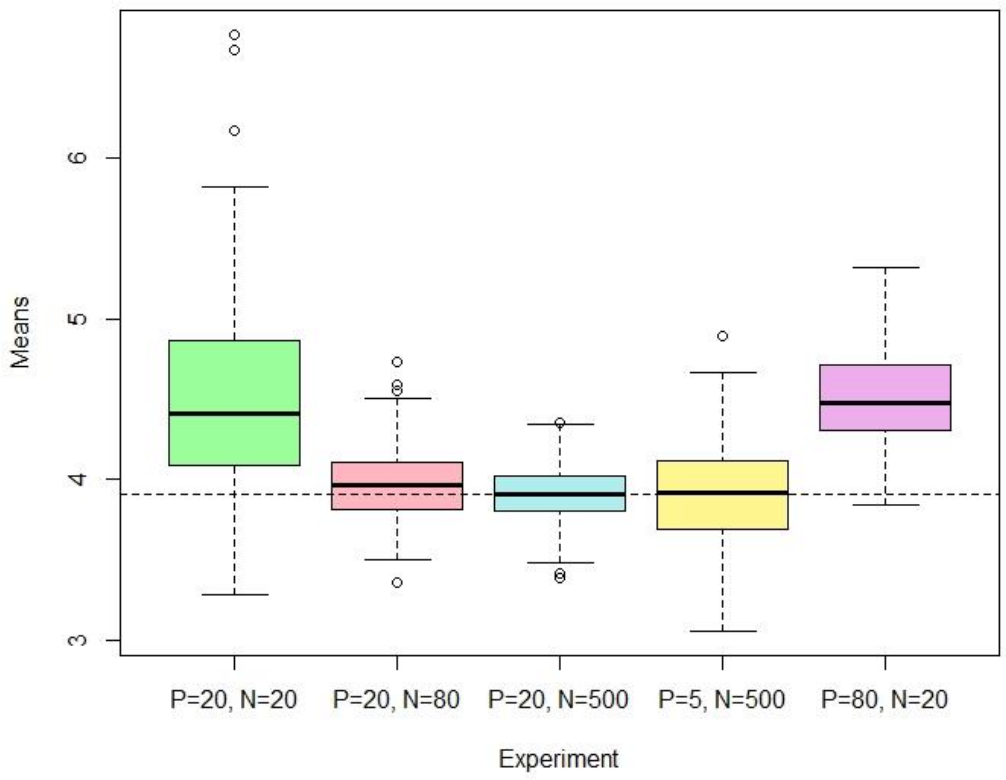
Note. P is the number of participants and *N* is the number of trials.

The importance of the number of trials was reflected visually, as seen in the boxplots of the CMLE drift rate estimates. Figure 3 depicts the original Faulkenberry et

al. (2018) drift rate of 3.91 as a dashed line, with the CMLE estimates depicted as a boxplot for each sub-experiment. While the number of participants could vary between 5 and 20 subjects, having at least 80 trials per participant resulted in boxplots centered close to the original drift rate. The two experiments with 500 trials both had means that aligned perfectly with the true parameter averages. Contrastingly, the two experiments with 20 trials had mean drift rates well above the mean. The only notable difference between the two experiments with 20 trials was that the experiment with 20 participants had a larger interquartile range than the experiment with 80 participants.

Figure 3

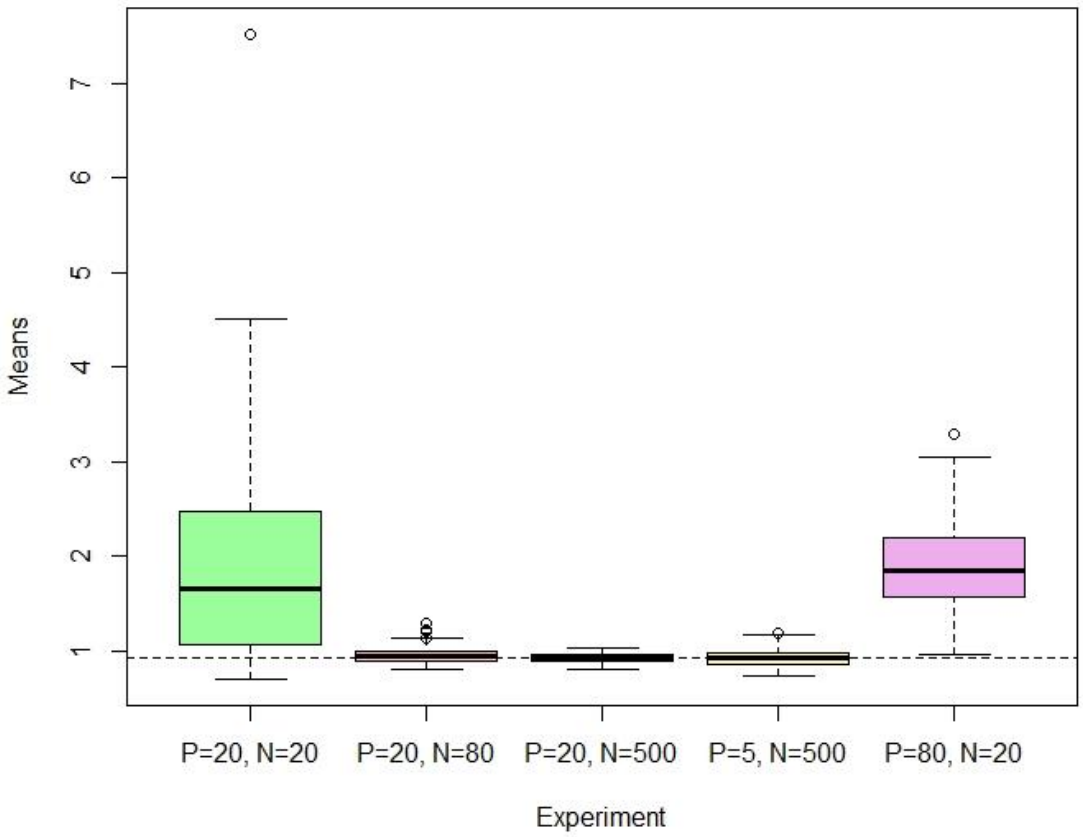
Boxplots for CMLE Estimated Mean Drift Rates



The impact of the number of trials was even more pronounced in the boxplots for the mean response thresholds, as seen in Figure 4. The average for the CMLE response threshold estimates directly aligned with the average for the true parameters, with very little variability in scores. Similar to the boxplots for drift rate, the response threshold means for the experiments with 20 trials were well above the original parameter, with the sub-experiment with 20 participants having a slightly larger interquartile range.

Figure 4

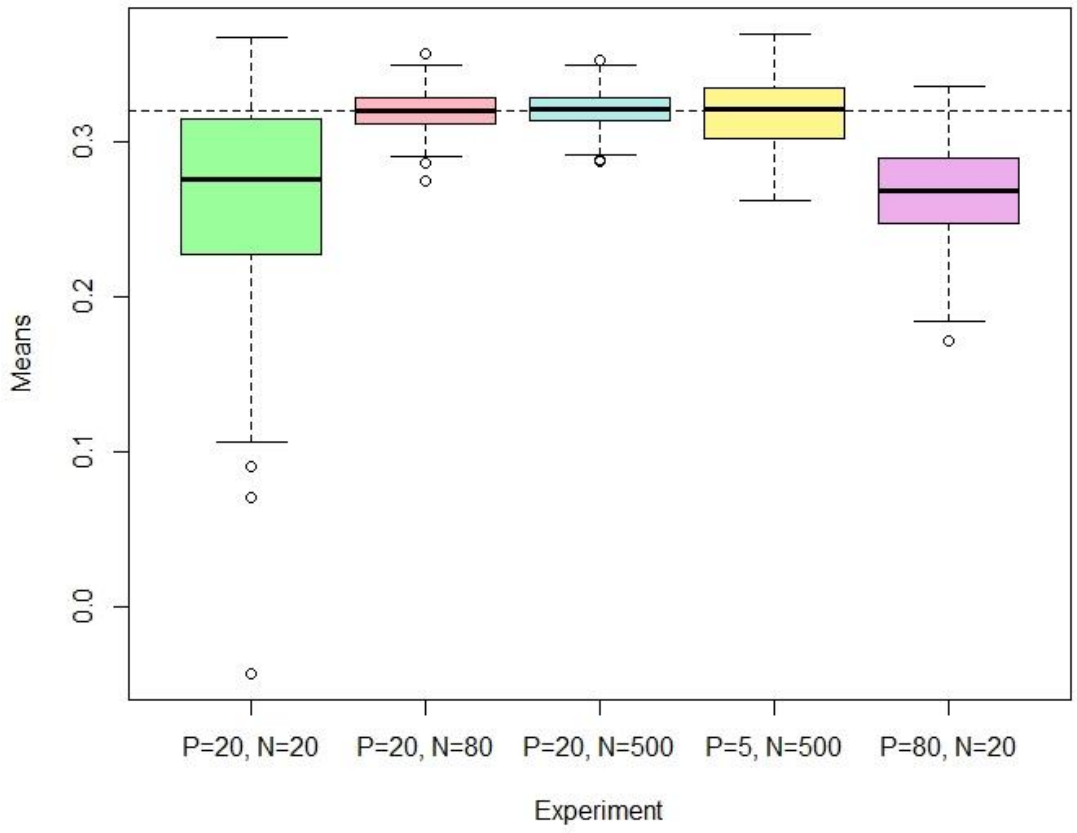
Boxplots for CMLE Estimated Mean Response Thresholds



Similar to what was seen with the CMLE estimates for drift rate and response threshold, Figure 5 shows that the experiments with more than 80 trials had CMLE averages that aligned with the true parameter average. While the two experiments with 20 trials had CMLE averages above the true parameter averages for both drift rate and response threshold, these two experiments had CMLE estimates below the true parameters for shift. Despite the opposite direction, the trend of a larger interquartile range for the lower participant count continued.

Figure 5

Boxplots for CMLE Estimated Mean Shifts



The first method of mathematically evaluating the differences between CMLE estimates and the true parameter estimates was RMSD. The average RMSD values for all three shifted-Wald parameters are listed for each sub-experiment in Table 2. Just as was seen in the visual display of the means, the sub-experiments with more trials per participant had substantially lower RMSD values, implying a higher accuracy, than the sub-experiments with 20 trials. Interestingly, between the sub-experiments with 20 trials, the experiment with only 20 participants was more accurate than the experiment with 80 participants for all three parameters.

Table 2

Root-Mean Squared Deviations Between the True and Estimated Parameters

Experiment	γ	α	θ
P = 20, N = 20	2.42	3.47	0.24
P = 20, N = 80	0.66	0.30	0.04
P = 20, N = 500	0.24	0.09	0.01
P = 5, N = 500	0.23	0.09	0.01
P = 80, N = 20	2.65	4.17	0.27

Note. Values given in seconds.

The mean bias was collected for all 200 samples in each experiment, and the average mean bias with standard deviations are displayed in Table 3. The sub-experiments with more trials had incredibly low mean biases, indicating an accuracy of the CMLE estimates. In fact, for the experiments with 80 or 500 trials, the average mean biases for shift centered at zero, implying that there was almost no bias. For both of the experiments with 20 trials, CMLE systematically overestimated drift rate and response threshold but underestimated shift, matching the results seen in Figures 3, 4, and 5.

Similar to the RMSD results for the experiments with 20 trials, the sub-experiment with 20 participants was slightly more accurate than the experiment with 80 participants.

Table 3

Mean Bias (B) and Standard Deviation (SD) of the Estimated Parameters

Experiment	γ		α		θ	
	B	*SD*	*B*	*SD*	*B*	*SD*
P = 20, N = 20	0.58	0.58	0.96	0.98	-0.05	0.06
P = 20, N = 80	0.06	0.15	0.03	0.07	0.00	0.01
P = 20, N = 500	0.01	0.05	0.00	0.02	0.00	0.00
P = 5, N = 500	0.02	0.10	0.01	0.04	0.00	0.01
P = 80, N = 20	0.61	0.31	0.98	0.48	-0.05	0.04

The correlations between the CMLE estimates and the true parameters were calculated, and the average correlation for each sub-experiment is listed in Table 4. For all experiments, response threshold had the weakest correlation of all three parameters. The two sub-experiments with 500 trials had the highest correlations, while the experiment with 80 trials had more mid-level correlations. Both sub-experiments with 20 trials had very weak correlations between the CMLE estimated parameters and the true parameters, especially for the response thresholds. Matching the results seen with both RMSD and mean bias of these two sub-experiments, the experiment with 20 participants had marginally higher correlations than the experiment with 80 participants.

Table 4

Correlations Between the True and Estimated Parameter Values

Experiment	γ	α	θ
P = 20, N = 20	0.36	0.16	0.30
P = 20, N = 80	0.73	0.53	0.80
P = 20, N = 500	0.94	0.87	0.97
P = 5, N = 500	0.93	0.86	0.95
P = 80, N = 20	0.32	0.12	0.21

To illustrate the correlations listed in Table 4, correlation plots were generated from one sample of each of the sub-experiments. Figure 6 depicts a sample of 20 participants with 20 trials each. The average correlation for all 200 samples is listed for each parameter. As there are 20 participants, there are 20 data points comparing the true value for each parameter against the CMLE value. Despite the fact that 20 data points exist for each plot, not all 20 are visible on the graphs. For this experiment, the correlation was so weak that not all CMLE parameter values fell within the normal range, especially for drift rate and response threshold.

Figure 6

Correlation of a Sample Where P = 20 and N = 20

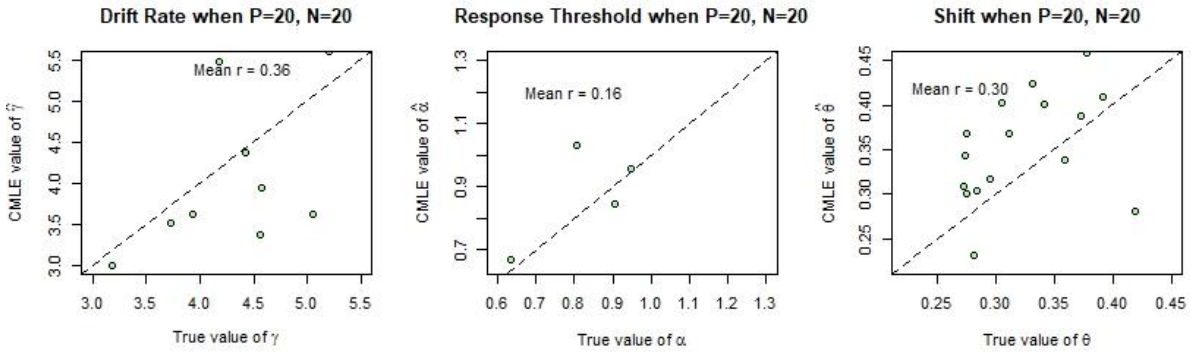
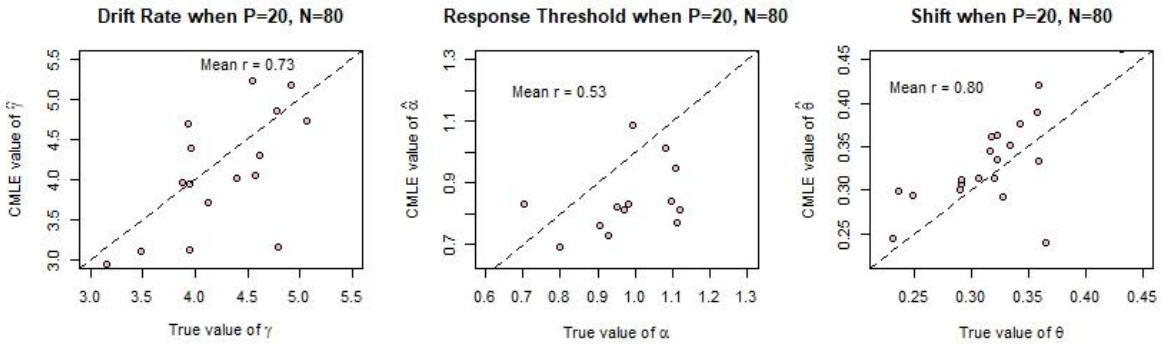


Figure 7 illustrates the correlation for a sample of 20 participants with 80 trials each. While the correlations were still not perfect, increasing the number of trials vastly improved the correlation between the CMLE values and the true values.

Figure 7

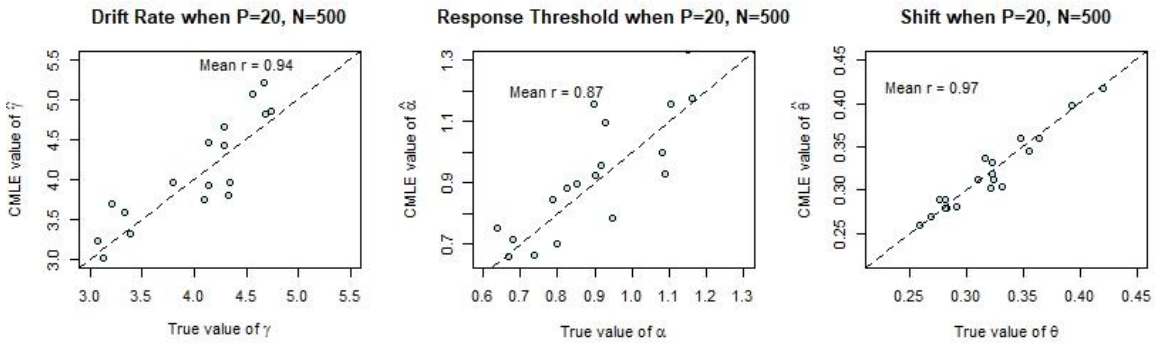
Correlation of a Sample Where P = 20 and N = 80



In Figure 8, the number of trials in the sample was raised to 500, and the correlations between the true values and CMLE values were plotted. Shift had the highest correlation between values, followed by drift rate and then response threshold. This illustrates how CMLE accurately estimated the parameter values when there were 500 trials.

Figure 8

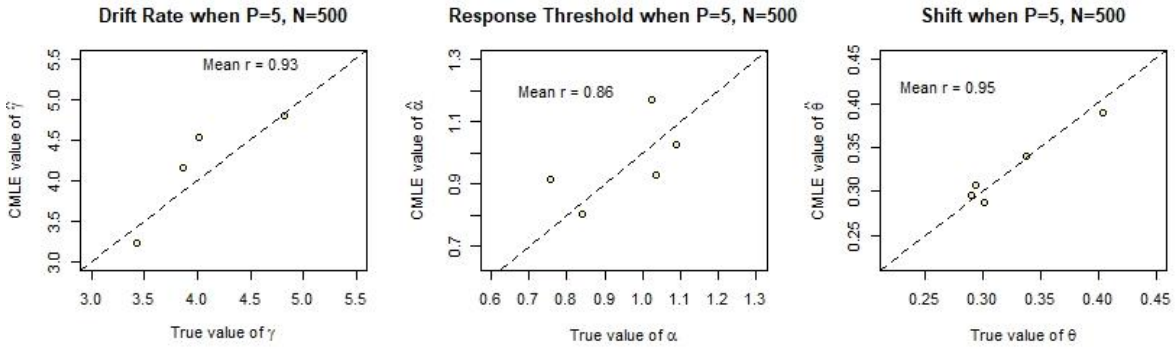
Correlation of a Sample Where P = 20 and N = 500



For the next sub-experiment, the number of trials remained at 500, but the number of participants was decreased to 5. An example sample from this experiment is depicted in Figure 9. Despite there being one-fourth the number of participants seen in Figure 8, the correlations for this experiment were extremely close to the experiment with 500 trials for 20 participants. As seen in the previous experiments, the correlation between true and estimated values was highest for shift and lowest for response threshold. This illustrates that even with very few participants, CMLE can accurately estimate the parameters, as long as there is a large number of trials per participant.

Figure 9

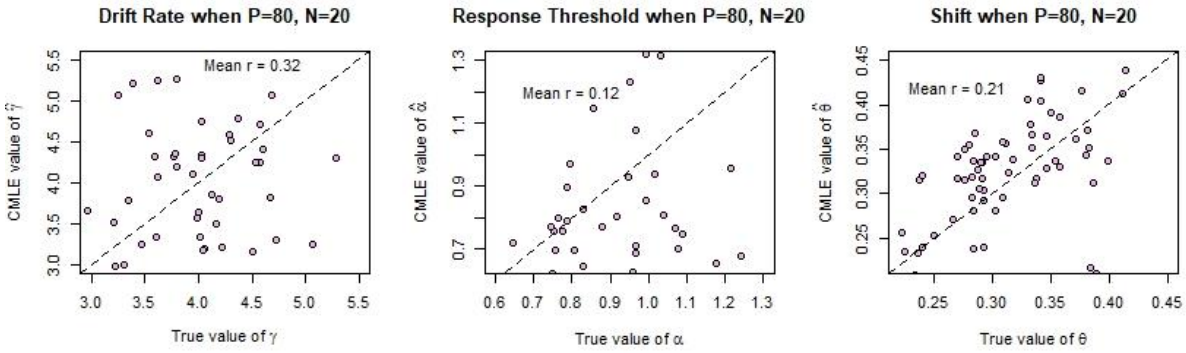
Correlation of a Sample Where P = 5 and N = 500



In the final sub-experiment, there were 80 participants with 20 trials each, and a sample of correlations from this experiment is depicted in Figure 10. Despite being the experiment with the highest number of participants (having four times as many participants as the first three sub-experiments and 16 times participants as the fourth sub-experiment), this sub-experiment had the weakest correlations. Only a fraction of the 80 data points are visible within the normal ranges of scores for each parameter. This indicates that CMLE could not estimate the shifted-Wald parameter values with much accuracy if there were only 20 trials per participant, whether the number of participants was increased or not.

Figure 10

Correlation of a Sample Where P = 80 and N = 20



CHAPTER IV

DISCUSSION

In this study I predicted that CMLE would perform well in shifted-Wald models, with low RMSD values, low mean biases, and high correlations. Overall, CMLE met these expectations, with RMSD values ranging between 0.01 and 4.17, mean biases ranging between -0.05 and 0.98, and some correlations as high as 0.97 (though there were some very low correlations). Increasing the number of trials per participant lowered RMSD values and mean biases, while raising correlations. Notably, the number of participants did not have the same impact. For the two experiments with 500 trials, having 20 participants instead of 5 had minimal to no change on RMSD, mean bias, or correlations. Surprisingly, for the two trials with 20 trials, increasing the number of participants from 20 to 80 actually resulted in slightly higher RMSD valuess and mean biases, and slightly lower correlations.

To evaluate how well CMLE estimated parameters in shifted-Wald models compared to when it was used with another model, such as the ex-Gaussian model, I juxtaposed my results against those of Farrell and Ludwig (2008). In terms of RMSD and mean bias, CMLE performed better when used with ex-Gaussian models. In the sub-experiment with 20 participants and 20 trials, Farrell and Ludwig (2008) found RMSD values of 0.0434, 0.0297, and 0.0478 for the ex-Gaussian components of mean, standard deviation, and the tail component, respectively. In comparison, for the shifted-Wald model's 20 participants with 20 trials sub-experiment, I found RMSD values of 2.42, 3.47, and 0.24 for drift rate, response threshold, and shift, giving results close to the ideal RMSD value of zero, but not as close as seen in ex-Gaussian models. For the sub-

experiment with 20 participants and 500 trials, Farrell and Ludwig (2008) found RMSD values of 0.0062, 0.0046, and 0.0075 for mean, standard deviation, and the tail component, while my shifted-Wald RMSD values were 0.24, 0.09, and 0.01 for drift rate, response threshold, and shift, again being outperformed by the ex-Gaussian application.

The trend in the comparison between the ex-Gaussian model of Farrell and Ludwig (2008) and my shifted-Wald model was seen in the mean bias results as well. For the 20 participants and 20 trials sub-experiment, Farrell and Ludwig (2008) found a mean bias of 0.0127 for the ex-Gaussian mean (*SD* = 0.1997), a mean bias of -0.0047 for the standard deviation (*SD* = -0.0305), and a mean bias of -0.0123 for the tail component (*SD* = 0.0508). For the same sub-experiment, I found a mean bias of 0.58 for drift rate (*SD* = 0.58), a mean bias of 0.96 for response threshold (*SD* = 0.98), and a mean bias of -0.5 for shift (*SD* = 0.06). While in the sub-experiment with 20 participants and 20 trials, CMLE performed better in the ex-Gaussian parameter estimation than in the shifted-Wald model parameter estimation, this difference decreased when the number of trials was increased to 500, especially when there were only 5 participants. In this experiment with 5 participants and 500 trials, Farrell and Ludwig (2008) reported mean biases of 0.0002 (*SD* = 0.182), -0.002 (*SD* = 0.0106), and -0.001 (*SD* = 0.0233) for mean, standard deviation, and the tail component, respectively. These are rather similar to my mean bias results of 0.02 (*SD* = 0.10), 0.1 (*SD* = 0.04), and 0.00 (*SD* = 0.01) for drift, response threshold, and shift, respectively. For this sub-experiment, the main difference was that my values had less variation for two variables. These results may indicate that not only does increasing trials improve the accuracy within a model, but also that a higher number of trials also makes CMLE better applicable to other models as well.

While CMLE performed in the shifted-Wald models relatively similar to ex-Gaussian models in terms of RMSD and mean bias, the correlations results were more drastically different, especially when there was a small number of trials. For the sub-experiment with 20 participants and 500 trials, Farrell and Ludwig (2008) found correlations of 1, 0.91, and 0.95 for the mean, standard deviation, and tail components. These are relatively similar to my shifted-Wald model correlations of 0.94, 0.87, and 0.97 for drift rate, response threshold, and shift. The experiments with fewer trials were a different story, especially for the sub-experiment with 80 participants and 20 trials. In this sub-experiment Farrell and Ludwig (2008) found correlations of 0.98, 0.33, and 0.46 for the mean, standard deviation, and tail components, while my shifted-Wald model had correlations of 0.32, 0.12, and 0.21 for drift rate, response threshold, and shift. For this experiment, CMLE performed nearly three times better in the ex-Gaussian model than in the shifted-Wald model. This may indicate that CMLE may not be as well-equipped for experiments with small trial sizes as would be expected from the RMSD and mean bias results.

The results of this study highlight an important debate within the world of cognitive psychometrics: whether the number of trials is more important than number of participants. Smith and Little (2018) compared designs with small or large numbers of participants in individual- and group- level analysis and found that a very large increase in the number of participants was needed to make the larger study designs comparable to the small, individual-focused designs. In addition, the larger study designs resulted in the obscuring of individual differences and effect sizes. The experiment designs with fewer participants consistently outperformed the larger designs, proving much more accurate in

parameter recovery. Smith and Little (2018) also argued that smaller designs are more applicable to the field, in terms of both available resources (as it can be difficult and expensive to involve a large number of participants) and the underlying concepts. From this standpoint, each individual is treated as an experiment unto themselves, and the focus becomes studying the psychometric concept in question over multiple trials, splitting differences in individuals into a separate question.

The argument for a smaller number of participants but larger number of trials is supported by the results of this study. When comparing the results of the sub-experiment with 20 participants with 20 trials and the sub-experiment with 80 participants with 20 trials, the experiment with the lower number of participants had lower RMSD values, lower mean biases, and higher correlations. When looking at the boxplots between these two experiments, they arrived at the same mean for all three shifted-Wald parameters. The only difference between these two experiments was that the experiment with 80 participants had smaller interquartile ranges, possibly indicating a higher reliability, but not a higher validity. As both experiments arrived at nearly identical means for all three parameters, it is possible that both sub-experiments with 20 trials were getting caught on a local maximum. If this is the case, increasing the number of participants had no impact on preventing the local maximum problem, while increasing the number of trials did successfully avoid any local maxima and estimated the correct parameter. This debate about the importance of number of participants versus the number of trials opens an entire realm of possible experiments to study, both with CMLE being applied to other models, as well as other estimation methods being used with shifted-Wald.

Another question that can arise from the results with only 20 trials revolves around the accuracy indicated by the RMSD values and mean bias, but the low correlations. Is there an error in CMLE estimating parameters from shifted-Wald models or is there a bias in Pearson's correlation itself? Taking this a step further, is there a "goal" trial size that eradicates the contrasting results between the correlations and the values for RMSD and mean bias? The reasoning behind these low correlations is an area of possible future research in itself.

Finally, this project illustrated that CMLE adequately estimates parameters for shifted-Wald models, especially when there is a larger number of trials. When comparing the different sub-experiments, it is evident that the number of trials is crucial for accurately estimating shifted-Wald parameters. So now what? It is not enough to say that CMLE estimated the parameters well, without comparing it to any other estimation method. After all, it may be that while CMLE estimated parameters for shifted-Wald models just as well as it did for ex-Gaussian models, CMLE may pale in comparison to other estimation methods. So just as Farrell and Ludwig (2008) compared CMLE to other methods, a great next step in the research for this area would be to evaluate if CMLE is the best tool for the job, or if other estimation methods will outshine CMLE in shifted-Wald models.

REFERENCES

Agresti, A. (1989). Tutorial on modeling ordered categorical response data. *Psychological Bulletin, 105*(2), 290–301. https://doi.org/10.1037/0033-2909.105.2.290

Alexandrowicz, R. W., & Gula, B. (2020). Comparing eight parameter estimation methods for the Ratcliff diffusion model using free software. *Frontiers in Psychology, 11*, 1-22. https://doi.org /10.3389/fpsyg.2020.484737

Anders, R., Alario, F.-X., & Van Maanen, L. (2016). The shifted Wald distribution for response time data analysis. *Psychological Methods*, *21*(3), 309-327. https://doi.org /10.1037/met0000066.supp

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc, 36*(4), 742–756. https://doi.org /10.3758/bf03206555

Etz, A. (2018). Introduction to the concept of likelihood and its applications. *Advances in Methods and Practices in Psychological Science*, *1*(1), 60-69. https://doi.org/10.1177/2515245917744314

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*(6), 1209-1217. https://doi.org /10.3758/PBR.15.6.1209

Faulkenberry, T. J. (2017). A single-boundary accumulator model of response times in an addition verification task. *Frontiers in Psychology*, *8*(1,225), 1-12. https://doi.org/10.3389/fpsyg.2017.01225

Faulkenberry, T. J. (2021, June 8). *MathPsych – Lecture 3* [Video]. YouTube. https://www. youtube.com/watch?v=TiW3M21H9U4

Faulkenberry, T. J., Vick, A. D., & Bowman, K. A. (2018). A shifted Wald decomposition of the numerical size-congruity effect: Support for a late interaction account. *Polish Psychology Bulletin*, *49*(4), 391-397. doi.org/10.24425/119507

Hübner, R., & Pelzer, T. (2020). Improving parameter recovery for conflict drift-diffusion models. *Behavior Research Methods*, *52*(5), 1848-1866. https://doi. org/ 10.3758/s13428-020-01366-8

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16*(5), 798-817. https://doi.org/10.3758/pbr.16.5.798

Miller, R., Scherbaum, S., Heck, D. W., Goschke, T., & Enge, S. (2018). On the relation between the (censored) shifted Wald and the Wiener distribution as measurement models for choice response times. *Applied Psychological Measurement, 42*(2), 116-135. https://doi.org /10.1177/0146621617710465

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90-100. https://doi.org /10.1016/S0022-2496(02)00028-7

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*(3), 190–214.

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers, 33*(4), 457-469. https://doi.org/10.3758/bf03195403

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-*N* design. *Psychonomic Bulletin and Review*, *25*, 2083–2101. https://doi.org/ 10.3758/ s13423-018-1451-8

Steingroever, H., Wabersich, D., & Wagenmakers, E.-J. (2021). Modeling across-trial variability in the Wald drift rate parameter. *Behavior Research Methods, 53*(3), 1060–1076. https://doi.org /10.3758/s13428-020-01448-7

van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology*, *53*(6), 463–473. https://doi.org /10.1016/j.jmp.2009.09.004

Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3. https://doi.org /10.3758/bf03194023

APPENDIX

R SCRIPT

```
# various utility functions for parameter recovery routines


# density function for Wald (unshifted)
dwald = function(x,gamma,alpha){
  return((alpha/(sqrt(2*pi*x^3)))*exp(-(alpha-gamma*x)^2/(2*x)))
}

# function to generate random shifted Wald data
# adapted from pp. 79-80, Dagpunar, J. (1988). Principles of Random
Variate Generation. Clarendon Press, Oxford.
# code modified from Heathcote (2004)
rwald = function(n, gamma, alpha, theta) {
  y2 = rchisq(n,1)
  y2onm = y2/gamma
  u = runif(n)
  r1 = (2*alpha + y2onm - sqrt(y2onm*(4*alpha+y2onm)))/(2*gamma)
  r2 = (alpha/gamma)^2/r1
  ifelse(u < alpha/(alpha+gamma*r1), theta+r1, theta+r2)
}

# basic function that returns Root Mean Squared Error
rmse = function(error){
  sqrt(mean(error^2))
}

# negative log likelihood for shifted Wald
nll.wald = function(par,dat){
  return(-sum(log(dwald(dat-par[3], gamma=par[1], alpha=par[2]))))
}

# Start point estimate for SW, based on first two moments
# assumes s = p*min(x), where x is a data vector
# from Heathcote (2004)
waldinit = function(x, p = 0.9) {
  theta = p*min(x)
  x = x - theta
  gamma = sqrt(mean(x)/var(x))
  alpha = gamma*mean(x)
```

```
  return(c(gamma,alpha,theta))
}
#Running simulations for Fit Functions

# define simulation parameters

nSub = 20
nTrials = 20

# parent distribution means and sds (from F, Vick, Bowman, 2018)
G = 3.91 # gamma
G.sd = 0.70
A = 0.92 # alpha
A.sd = 0.17
H = 0.32 # theta
H.sd = 0.05

# each simulation run starts HERE
# build simulated RT matrix
# rows = subjects, columns = trials
rts = matrix(0, nrow=nSub, ncol=nTrials)

# build matrix to store "target" parameters
# rows = subjects, columns = gamma, alpha, theta
targets = matrix(0, nrow=nSub, ncol=3)

# randomly draw target shifted Wald parameters from parent distribution
# (one unique parameter value for each subject)
for (i in 1:nSub){
  targets[i,1] = rnorm(1, mean=G, sd=G.sd) # draw random gamma
  targets[i,2] = rnorm(1, mean=A, sd=A.sd) # draw random alpha
  targets[i,3] = rnorm(1, mean=H, sd=H.sd) # draw random theta
}

# from these target SW values, generate the distribution of observed RTs
for each subject

for (i in 1:nSub){
  rts[i,] = rwald(nTrials,
                  gamma = targets[i,1],
                  alpha = targets[i,2],
```

```
                    theta = targets[i,3])
}

# now let's use CMLE to fit shifted Wald distribution to each row of
observed RTs
# first, we'll store these estimates in a matrix
predictions = matrix(0, nrow=nSub, ncol=3)

# now do the fits (one per subject)
for (i in 1:nSub){
  fit = optim(waldinit(rts[i,]), nll.wald, dat=rts[i,])
  predictions[i,] = fit$par
}
```