2006

ANALYSIS OF REACTION TIME EXPERIMENT DATA USING POINT PROCESS TECHNIQUES

Jennifer Lea Asimit
Western University

ANALYSIS OF REACTION TIME EXPERIMENT DATA USING POINT PROCESS TECHNIQUES

(Spine title: Analysis of Reaction Time Experiment Data)

(Thesis format: Monograph)

by

Jennifer Lea Asimit

Graduate Program
in
Statistics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
FACULTY OF GRADUATE STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. W. John Braun	Dr. R.J. Kulperger
Supervisory Committee	Dr. Duncan Murdoch
	Dr. John Koval
	Dr. Bruce Smith

The thesis by

Jennifer Lea Asimit

entitled:

Analysis of Reaction Time Experiment Data Using Point Process Techniques

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date_____

_____
Chair of the Thesis Examination Board

Abstract

Point process data with a latent or hidden component arise in a variety of research fields. Examples include reaction time, forest fires, seismology, and transactional data. We develop statistical methodology that can be applied to point process data with a latent component. The methodology is applied to data from a number of reaction time (RT) experiments.

Reaction time experiments have been of interest to psychophysicists for more than a century. A reason for this interest lies in the fact that the time taken to perform a task indicates the complexity of the operations occurring in the brain. We study three types of visual-motor RT experiments that increase in complexity.

The main contributions of my thesis are the development of two types of models that can be used in the analysis of point process data, in particular RT data. For each of the three RT experiments, we develop a parametric model to help provide a foundation for understanding the behavior of nonparametric intensity estimates. We also study threshold models and introduce different variants for the three RT experiments. The parameters in a threshold model have direct biological interpretations regarding inferences on the eye-brain-hand system.

Additional contributions include the use of nonlinear regression in parameter esti-

mation for our simple RT parametric model. This estimation method is useful when it is of interest to obtain a single set of parameter estimates for data sets obtained using different flash rates. We also provide a derivation of a covariance that is necessary for a hypothesis test by Asimit and Braun (2005).

Keywords: *point process, hidden processes, nonparametric estimation, intensity function, parametric modelling, Brillinger mixing, kernel smoothing, threshold model, generalized linear model, ordinal logit model, integrate-and-fire model.*

Acknowledgements

First, I would like to thank my supervisor Dr. John Braun for his constant support and encouragement. While providing guidance and useful suggestions, he helped me gain confidence to think more independently.

I also wish to express my gratitude to Dr. Bill Simpson of the SMART Section, DRDC Toronto for some helpful discussions, and for providing the RT experimental data.

I would like to express thanks to my parents Tony and Rose Prokop for their encouragement to achieve my goals.

Finally, I would like to thank my husband Vali whose love and support helped me throughout the duration of this thesis. Our son Stephan is thanked for many smiles that relieved times of stress.

This research was funded by the Natural Sciences and Engineering Research Council of Canada, to whom I am grateful for four years of graduate funding.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Researchers in visual psychophysics study the brain mechanisms underlying vision by presenting visual stimuli and obtaining behavioral responses from an observer (see Cornsweet (1970)). In a reaction time (RT) experiment, the experimenter presents stimuli and an observer hits a button in response to the stimuli. RT has been extensively studied, and is of interest to psychophysicists because the time taken to perform a visual-motor task indicates the complexity of the operations taking place in the brain, for example. We consider three variations of RT experiments: simple, go-no go, and choice.

Simple RT is concerned with detecting the occurrence of a stimulus, without identifying its type. In a simple RT experiment flashes are presented to a subject who responds as quickly as possible by pressing a button. The times of the flashes and responses are recorded, but the correspondence between the two sequences is unknown. In a go-no go RT experiment two flash types are presented and the observer is instructed to respond to only one type, by pressing a button. Thus, the observer needs to distinguish between the stimulus types presented, and respond only to a certain type of stimulus. Choice RT involves the additional task of responding differently to different stimuli; press button one when a type one stimulus is observed, and press button two when a type two stimulus is observed.

1.1 Overview

In Chapter 2 we review the history of reaction time experiments and the necessary background for the subsequent chapters. In that chapter we provide some of the fundamental point process definitions and properties, as well as a discussion on threshold models as latent variable models. The simple RT parametric model of Braun et al. (2003) is also reviewed there.

Simple RT data is studied in Chapter 3, which is a fundamental chapter that we often refer to in later chapters. A derivation of a covariance approximation necessary for the hypothesis test developed by Asimit and Braun (2005) is provided there. We also modify the parameter estimation method of Braun et al. (2003) and compare it with a nonlinear regression estimation method that we propose. Furthermore, we introduce a threshold model for simple RT, as well as diagnostics, and a variant that incorporates randomness due to experimental trials. The variables in a threshold model have direct biological interpretations.

We introduce two models for go-no go RT data in Chapter 4. First, we propose a parametric model, for which we derive expressions for certain point process intensity functions. Our parametric intensity estimates are used in conjunction with the corresponding nonparametric estimates to make inferences on the eye-brain-hand system. In the latter half of the chapter we introduce a threshold model for go-no go RT data. We fit both models to simulated and real data. Fitting the models to simulated data allows us to study the behaviour of our fitted models for data in which we know the relationship between the flashes and responses.

The objective of Chapter 5 is to introduce a parametric model for choice RT data. In that chapter, we derive intensity function expressions under our model, and fit our model to both simulated and real data.

Various threshold models for choice RT data are studied in Chapter 6. We first introduce an ordinal response threshold model that combines the information of both flash and response types. In order to extract more information from the data we also develop marginal threshold models for each response type. The inclusion of second-

order interaction terms in the marginal models allow us to make inferences regarding the effect of any flash interactions on the response types.

Finally, in Chapter 7 we summarize our results and discuss areas of further research.

Chapter 2

Literature Review and Preliminaries

The first reaction time experiment has been credited by Ribot (1900) to Helmholtz (1850), who followed an earlier recommendation of Dubois-Reymond (see Welford (1980)). Helmholtz designed a reaction time task in an attempt to measure the speed of neural transmission. In his experiment, the same type of response was made when various parts of the body were stimulated, so that impulses had to be sent along different nerve fibre lengths. Then, the velocities of neural transmission were calculated from the times taken over different measured lengths of fibres. Helmholtz concluded that the speed of neural transmission contributes to only a small portion of the total reaction time; central processes are a major component of reaction time.

One of the most important RT studies is the comparison of Donders (1868) of simple, go-no go, and choice reactions, from which he provided a method to measure mental processes (see Welford (1980)). Donders reasoned that since simple RT does not involve discrimination, and go-no go does, subtraction of simple RT from go-no go RT should yield the time for discrimination. Likewise, because discrimination is involved in both go-no go and choice RT, but a response only needs to be chosen for choice RT, the time spent in choosing a response can be found by subtracting go-no go RT from choice RT. Subsequent researchers suggested that there may be a time overlap of different processes, and that the analysis of complex elements cannot

always be decomposed into the analysis of simple elements. Eventually, the additive model of Donders lost favour.

In the traditional simple RT experiment the stimuli are presented on a trial-by-trial basis, so that the performance of a subject can be observed with ease. A complete summary is provided by Luce (1986). The stimuli are presented according to a certain distribution; a popular choice is the uniform distribution. The notion of a hazard function can be used to show why this is not a good experimental design. Recall that the hazard function $h(t)$ of a distribution gives the tendency for an event to occur at time t, and is defined by

$$h((t) = \frac{f(t)}{1 - F(t)},$$

where $f(\cdot)$ and $F(\cdot)$ are the density and cumulative distribution functions, respectively.

Upon presentation of a signal the hazard function is set to zero, and stays at that level for some amount of time after the response. Doing so prevents any overlapping of stimuli, and does not allow a subsequent stimulus to occur during the interval between a stimulus and its response. Thus, the stimuli are presented in discrete trials. In order to inform the subject when the hazard function becomes positive, so that it is time to be alert to respond to a stimulus, a *warning signal* is presented. The warning signal is usually a moderately intense stimulus of a modality different from the reaction stimulus; in the case that the stimuli are visual, the warning signal is usually auditory. The time interval between the warning signal and reaction stimulus is called the *foreperiod.* The RT is the time from the stimulus presentation to the response.

Many studies have been run using various distributions for the foreperiod lengths, with the uniform distribution being the most common (see Luce (1986)). A problem with using the uniform distribution for the foreperiods is that it has an increasing hazard function, which may help the subject predict when a stimulus will occur. That is, with a uniform delay distribution, the observer can anticipate the stimulus since as the observer waits, the chance of a stimulus occurring in the next instant increases.

In the case that the hazard function is constant, the stimulus can occur at any time after the warning signal, making it difficult to anticipate when to respond. The exponential distribution, which has density

$$f(x) = \lambda e^{-\lambda x},$$

is the only continuous distribution with a constant hazard function indicating that it is suitable for RT experiments. However, the exponential distribution is not without problems. Using an exponential distribution can occasionally lead to a very long time to wait for a stimulus to occur. It has been argued that after a few seconds of waiting the subject is unable to maintain a constant rate of anticipation. However, Luce (1986) cites empirical evidence that attention can be sustained for as long as nine seconds.

An exponential distribution for the time until a stimulus occurs solves the anticipation problem, but the traditional simple RT experiment has further deficiencies. Firstly, it is unrealistic to have the stimuli presented in discrete trials, and to follow a warning. In reality, signals that require a response occur without warning and at random times. For example, while driving a car, dangers appear and a driver must respond immediately by pressing the brake pedal. Another disadvantage of the traditional simple RT experiment is the impossibility of studying interaction effects between the stimuli, due to the discrete trials.

Presenting the flashes according to a point process and recording the times of the button presses resolves these issues. In a Poisson process with rate λ, the times between events are exponentially distributed with mean $1/\lambda$. Thus, due to the memoryless property of the exponential distribution, the times of events are unpredictable, and the anticipation problem is still addressed. Figure 2.1 displays a segment of simple RT data from an experimental run in which the flash process has rate $1/s$.

Evidence of the influence of the hazard function on RT was given by Simpson et al. (2000), who compared simple RT experiments in which flashes are presented as a uniform delay point process, and as a Poisson process. They found that when the stimuli have uniform delays the responses are very similar, and quicker, with speed

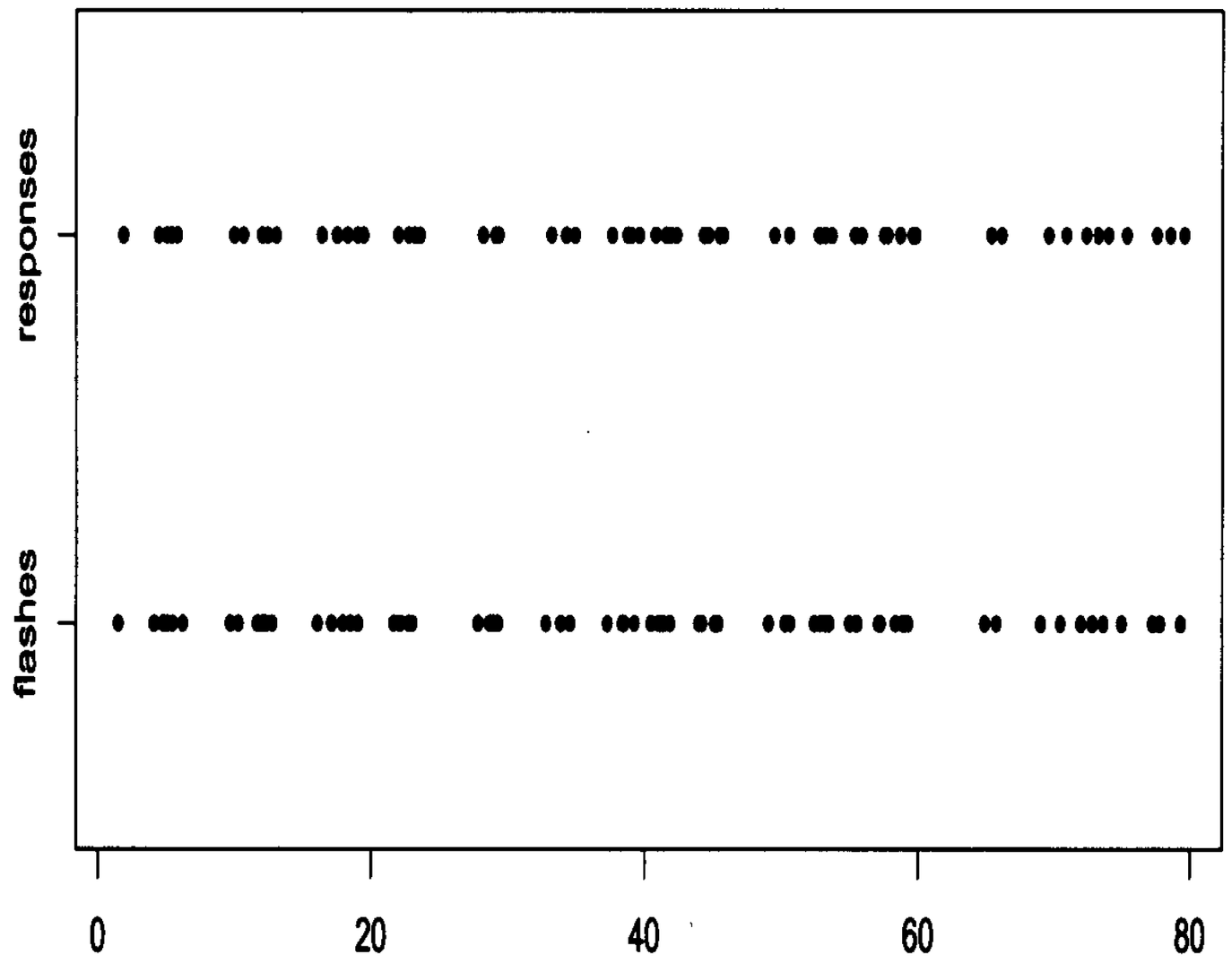


Figure 2.1: A display of RT data from the first 80 s of one experimental run with flashes presented at rate $1/s$. The complete experimental run has a length near 320 s, and consists of 300 flashes and 202 responses.

increasing as the wait for the stimulus increases. This led to the conclusion that any information contained in the hazard function is used by the observers. That is, the high level cognitive task of computing the hazard function is involved in simple RT. The modified simple RT experiment consists of presenting the flashes according to a Poisson process, and recording the times of the button presses.

A measure of RT is given by the location (in time) of the peak of the average impulse response of the eye-brain-hand system. A quantity of interest to psychophysicists is the irreducible minimum RT, which is defined as the fastest possible RT that is approached as stimulus strength (i.e. flash brightness) increases. In order to obtain the irreducible minimum RT, Simpson et al. (2000) measure RT as a function of

stimulus strength.

Simpson et al. (2000) tentatively inferred the presence of nonlinearity in the system; the presence of inhibition and facilitation among the flashes. However, they did not provide uncertainty measures. There is *inhibition* among flashes if a pair of flashes at certain times is associated with a decrease in the response rate. *Facilitation* is present when a pair of flashes at certain times is associated with an increase in the response rate.

We focus on visual RT experiments involving responses to flashes of light, but various types of reaction time experiments exist in the literature. Diederich (1995) studied a simple RT experiment in which the subject had to respond to three different types of sensory stimuli (visual, auditory, tactile) presented alone, in pairs, or as triples to a subject. The phenomenon of a decrease of RT when two or more stimuli are presented, compared to the RT when only one stimulus is presented (intersensory facilitation) was observed, and his objective was to explain this occurrence.

Visual motor RT has been explored by Genova et al. (2000) with the purpose of studying how the visual system provides input for maintaining equilibrium and control of motor behavior, when there are quick changes in the direction of the optic flow. In this context, an RT experiment involved responding when there was a change in the direction (left or right) and angle of motion of a vertically upwards moving random dot pattern. In a simple RT experiment the subjects press a button regardless of the direction change, while in a choice RT experiment a left or right button is pressed, corresponding to the change of direction perceived.

Kreegipuu et al. (2006) investigated reaction times for detecting a change in the color or contrast of an object, stationary or moving at constant velocity. They found that the reaction time to detect a change in colouration decreased as the speed of the moving object increased, indicating that the mechanism processing colour is aware of movement parameters.

In Section 2.1 we review the basic neurophysiology of visual perception in RT experiments, while some fundamental point process definitions and properties are given in Section 2.2. We discuss a point process system identification problem in Section 2.3,

where we also define associated point process parameters. In Section 2.4 we review the point process operations used in a parametric model for simple RT introduced by Braun et al. (2003), which we review in Section 2.5. We discuss nonparametric intensity estimation in Section 2.6. Threshold models as latent variable models are discussed in Section 2.7.

2.1 Neurophysiology of Visual Perception in RT Experiments

RT experiments involving responses to flashes of light are used by researchers in visual psychophysics to learn about the brain mechanisms underlying vision. A thorough account on the details of visual perception is given by Cornsweet (1970). We discuss some of the fundamentals.

The retina is an important component of visual perception. It contains five layers of cells, but the two most important layers are the photoreceptors and the ganglion cells (see Deutsch and Deutsch (1993)). The photoreceptors receive the light energy and transform it into a graded receptor potential, which is in turn converted into action potentials in ganglion cells; action potentials are the neuronal output of the receptor potential. There are two types of photoreceptors: cones and rods. Cones are responsible for colour vision and high resolution vision, while rods are used to see in dim light and control lower resolution, black-and-white night vision. In the reaction time experiments of interest to us, black and/or white flashes are presented against a grey background in a dark room, to a subject who has been adapted to the dark. Thus, our interest is in studying the rods and black-and-white night vision.

Images are formed on the retina, which contains millions of rods (see Cornsweet (1970)). Rhodopsin is a photopigment that is a mixture of retinal and protein molecules. After a flash of light is presented a number of rods in the eye retinas change shape (isomerize) causing a change in the electrical current produced by the rod. Some isomerizations are due to flashes and others occur spontaneously. The rod cur-

rent is proportional to the number of isomerizations, and when the current crosses a threshold level, a retinal ganglion cell is activated to generate an impulse that is transmitted down its axon to act upon other neurons in the brain, and the observer will decide that a flash was presented.

2.2 Point Processes

A point process on the real line may be used to model the occurrences of events. An excellent reference on point processes is Daley and Vere-Jones (2003); here we discuss some of the important properties and definitions. A point process on the line can be described in one of four equivalent ways:

- counting measures;
- nondecreasing integer-valued step functions;
- sequences of points;
- sequences of intervals.

We take an informal approach here. Let A be a point process with event times $\{A_1, A_2, \ldots, \}$, and counting measure N. For a Borel subset X of the real line, $N(X)$ denotes the number of times that events of A occur in the set X; that is

$$N(X) = \sharp\{i : A_i \in X\}.$$

When X is the half-open interval $(a, b]$ we follow the convention of using the more convenient notation $N(a, b]$, rather than $N((a, b])$. Likewise, the number of events in $(0, t]$ is denoted $N(t)$.

In the case where $N(-\infty, 0] = 0$ the equivalency of describing a point process by specifying the points A_i and specifying the function $N(t)$ is made clear by noting the following relation:

$$\inf\{t > 0 : N(t) \geq i\} \leq t \text{ if and only if } N(t) \geq i.$$

In the reaction time experiments we study, the stimulus process is a *homogeneous Poisson process* on the real line. The stationary Poisson process possesses the exponential increment property. That is, the times between events of a homogeneous Poisson process with rate λ are independent and exponentially distributed with mean $1/\lambda$.

2.2.1 Definitions

A point process has the property of *crude stationarity* (or simple stationarity) if the distribution of the number of points lying in an interval depends only on its length, and not its location, so that $\Pr\{N(x) = k\} = \Pr\{N(t, t+x] = k\},\ x > 0, k = 0, 1, \ldots$. This is a weaker property than stationarity.

Definition 2.2.1. A point process is *stationary* when the joint distribution of

$$\{N(A_1 + t), \ldots, N(A_r + t)\}$$

does not depend on t for all $r = 1, 2, \ldots$ and all bounded Borel subsets $A_1, \ldots, A_r$ of the real line.

Definition 2.2.2. A point process is *simple* when

$$\Pr\{N(\{t\}) = 0 \text{ or } 1 \text{ for all } t\} = 1.$$

The *intensity* λ of a (crudely) stationary point process is a measure of the rate, and is given (informally) by

$$\lambda = \lim_{h \downarrow 0} \frac{\Pr\{N(0, h] > 0\}}{h}. \tag{2.2.1}$$

For a simple point process with finite λ, the quantity λh gives an approximation of the probability that there is at least one point in an interval of length h. That is, we have

$$\Pr\{N(x, x + h] > 0\} = \lambda h + o(h).$$

The conditional intensity function $\mu(t)$ determines the probability structure of the point process uniquely (Daley and Vere-Jones (2003)), and is defined by

$$\mu(t) = \lim_{\Delta t \to 0} \frac{\Pr(N(t, t + \Delta t) > 0 | H_t)}{\Delta t},$$

where H_t is the history of the process until time t. It may be thought of as the frequency with which events are expected to occur around a particular point in time, conditional on the prior history of the point process, which may be multivariate. For the point process system (A, B), $\mu_B(t)$ is the conditional rate of the B process given the history of both the A and B processes.

Approximating the conditional intensity function is one of the problems of system identification, and is discussed in Section 2.3. This involves the estimation of certain point process parameters, which are defined and discussed in Section 2.3.1.

2.3 A Point Process System Identification Problem

A point process system has input and output processes that are both point processes. The identification problem for a point process system is to determine features of the system based on observations of inputs and output. In general, complete identification is impossible, so the objective is to determine average quantities or parameters that characterize the statistical properties of the system. The identification problem may also be described as determining an estimate of a finite dimensional parameter that characterizes the system behavior (see Brillinger (1975b)). Prior to giving some details of the problem, we provide the definitions of some point process parameters.

2.3.1 Point Process Parameters

Let (A, B) be a stationary bivariate point process on the real line, where A is an input process, and B is the corresponding output process. We denote the number of A events, respectively B events, in the interval X by $A(X)$, respectively $B(X)$. The following parameters describe stochastic point processes, and are as defined in Brillinger (1975b).

The mean density, or first-order intensity, of the B process, p_B satisfies

$$E[B(Y)] = p_B|Y|_L,$$

where $|Y|_L$ denotes the length of the interval Y. In the context of RT data, p_B is the response rate. The constant response rate implies that B is a first order stationary point process. Later we show that the response process is also second order stationary. Note that $p_B = E[B(0,1]]$.

A useful result is given by Korolyuk's Theorem (see, for instance, Daley and Vere-Jones (2003), p.47) and is stated below::

Theorem 2.3.1. *(Korolyuk's Theorem) For a crudely stationary simple point process,*

$$\lambda = E[N(0,1]], \text{ finite or infinite.}$$

From this theorem, it follows that $\lambda = E[B(0,1]] = p_B$ when B is a crudely stationary simple point process. Thus, using (2.2.1), the quantity $p_B h$ can be interpreted roughly as the probability of a B point occurring in an interval of length h. In the case of RT data p_B is the response rate.

The second-order intensity or second-order cross product density, $p_{AB}(u)$, satisfies

$$E[A(X)B(Y)] = \int_X \int_Y p_{AB}(y-x)\, dx\, dy.$$

An approximation to the probability that an A point occurs in an arbitrary interval of length h, and a B point occurs u time units later in a similar interval is given by $p_{AB}(u)h^2$. An analogous interpretation holds for $p_{BB}(u)$, which satisfies

$$E[B(Y_1)B(Y_2)] = \int_{Y_1} \int_{Y_2} p_{BB}(y_1 - y_2)\, dy_1\, dy_2 + p_B |Y_1 \bigcap Y_2|_L.$$

The third order intensity $p_{AAB}(u,v)$ satisfies

$$E[A(X_1)A(X_2)B(Y)] = \int_{X_1} \int_{X_2} \int_Y p_{AAB}(y-x_1, y-x_2) dx_1 dx_2 dy. \tag{2.3.1}$$

The function $p_{AAB}(u,v)$ provides a measure of the intensity with which the occurrence of a type B point coincides with the occurrence of two type A points u and v seconds earlier. Values of (u,v) that yield large $p_{AAB}(u,v)$ are those for which a B point is frequently found within u and/or v seconds from A point(s).

The conditional mean intensity, or AB cross-intensity function (CIF), is a first-order conditional rate function, and can be expressed as

$$m_{AB}(u) = p_{AB}(u)/p_A.$$

The quantity $m_{AB}(u)$ measures the average instantaneous rate of generating a B point at an interval near t, given that an A event occurs u time units earlier.

Cumulant densities give a measure of the dependence between processes. A cumulant density of order two is the cross-covariance density of (A, B), which gives a measure of the dependence as a function of the lag, and is defined as follows

$$q_{AB}(u) = p_{AB}(u) - p_A p_B. \tag{2.3.2}$$

As the dependence decreases between A and B points u units apart, $q_{AB}(u)$ approaches 0.

In the derivations of some of the intensity functions in later chapters we use the Dirac delta $\delta(\cdot)$, which satisfies

$$\int_{-\infty}^{\infty} \delta(x - x_0) f(x) dx = f(x_0).$$

2.3.2 System Identification

Identification of the point process system involves the construction of an expansion based on functions called "kernels", which is the point process analog of the Wiener-Volterra expansion (Brillinger (1975b), Brillinger (1976)). The kernels provide a complete characterization of the system, and can be expressed in terms of certain intensity functions. Brillinger's identification approach is used by Simpson et al. (2000) to characterize the time-varying rate of response $\mu_B(t)$, the conditional rate of response given the system history. We discuss system identification in the context of reaction time. The input process A consists of the flash times, and the times of the responses form the output process B.

The zeroth-order kernel a_0 is the mean response rate p_B, and when $\mu_B(t)$ is constant the simplest characterization is

$$\mu_B(t) = p_B.$$

However, if the observer is actually perceiving the flashes then the responses will not occur with the same frequency at all time instances. That is, the responses will be

more likely to occur a short time after a stimulus event. Introduction of the first-order kernel $a_1(t)$, which relates to the effect of a single flash, in the expansion for $\mu_B(t)$ reflects this dependence:

$$\mu_B(t) = a_0 + \int a_1(t-u)dA(u) = a_0 + \sum a_1(t - A_j),$$

where A_j is the time of the jth stimulus. The first order kernel is also called the impulse response, and is the best linear predictor of the average change of the instantaneous response rate at time t in a response train B, when a single stimulus occurs at time $t-u$ in a stimulus train A. Brillinger (1975b) has shown that when the input A is a Poisson process the first order kernel is $a_1(t) = \frac{p_{AB}(t)}{p_A} - p_B$. The lag at which the peak of $a_1(t)$ (or $p_{AB}(u)$) occurs is taken as a measure of the RT (Simpson et al. (2000)).

When two stimuli are presented close together in time there may be an inhibitive or facilitative effect on the response rate, indicating a nonlinear system. A second-order kernel $a_2(u,v)$ takes into account the effect of pairs of stimulus events on the response rate. When the second-order is included in the expansion we have:

$$\mu_B(t) = a_0 + \sum a_1(t - A_j) + \sum\sum_{k \neq j} a_2(t - A_j, t - A_k),$$

where A_j and A_k are two stimulus times. When the input is a Poisson process the second-order kernel is $a_2(u,v) = p_{AAB}(u,v) - p_{AB}(u) - p_{AB}(v) + p_B$.

The expansion may also involve higher order kernels, but (nonparametric) estimates of such kernels involve a large amount of data, and have a large variance (see Simpson et al. (2000)).

2.4 Point Process Operations

In order to form a parametric model for a point process version of an RT experiment, certain point process operations are required. Some operations of interest in our modelling are superposition, thinning, and random translation. Details of these operations follow, where we assume that multiple points do not occur at a single

time instance. For a point process A, the number of events up to time t is denoted by $N_A(t)$. Details of the operations are provided by Reiss (1993) and Stoyan et al. (1987).

Superposition is the sum of a finite or countable number of point processes. For example, the superposition of point processes A_1 and A_2 is $A = A_1 \bigcup A_2$, and we have $N_A(t) = N_{A_1}(t) + N_{A_2}(t)$. When A_1 and A_2 are independent homogeneous Poisson processes with respective rates λ_1, λ_2, the resulting superposed process is a Poisson process with rate $\lambda_1 + \lambda_2$.

Thinning is an operation that deletes points of a point process according to some definite rule. For the simplest thinning, called *p-thinning*, each point is deleted with probability $1 - p$ independently of the times and possible deletions of other points. More generally, the thinning probability can depend on the time of the point, so that a point at time t is deleted with probability $1 - p(t)$, where $p(\cdot)$ is a deterministic function, and deletion does not depend on the times or deletions of other points; this operation is referred to as $p(x)$-*thinning*. Both p-thinning and $p(x)$-*thinning* belong to the class of *independent thinnings*. Alternatively, thinning could depend on the configuration of the points in the point process - such thinnings belong to the class of *dependent thinnings*. A dependent thinning is employed in the parametric model for simple RT of Braun et al. (2003), described in Section 2.5.

Random translation is a special case of the *clustering* operation, which replaces every point x of a point process A by a cluster A_x of points. The cluster point process is the union of all of the clusters. In the case of random translation the size of each cluster is 1, and the ith cluster A_{x_i} is of the form $x_i + V_i$, where the V_i are random variables with density $f_i(\cdot)$. In the parametric model of Braun et al. (2003) the V_i are assumed to be independently and identically distributed with mean μ, variance σ^2, and common density $f(\cdot)$.

Thinning can also be viewed as a special case of the clustering operation in which each cluster A_x is either the singleton set $\{x\}$ or the empty set.

2.5 A Parametric Simple RT Model

We now review a parametric model for simple RT that was developed by Braun et al. (2003). In their parametric model the relationship between the flash process and response process in a simple RT experiment is modelled by the point process operations superposition, thinning and random translation. These operations are defined in Section 2.4.

Flashes are presented according to a homogeneous Poisson process with rate p_A per second. Therefore, the lengths of intervals between flashes are independent and exponentially distributed with mean $1/p_A$. The sequence $A = \{A_1, A_2, \ldots\}$ denotes the set of flash times on the real line, while the number of flashes in a time interval X is denoted $A(X)$. Similarly, $B = \{B_1, B_2, \ldots\}$ represents the set of response times, and $B(X)$ is the number of responses in time interval X. All flash and response times are reported in seconds. The flash and response processes can be thought to have been observed by the experimenter on an interval $[0, T]$. Note that A is taken to be a stationary Poisson process, and B is assumed to be a simple stationary point process.

Superposition is used to model a noisy flash process; it takes into account the possibility of mistakenly perceiving a flash, so that there is a response in the absence of a flash. The noise process is modelled as an independent Poisson process with rate p_S, and superposed onto the flash process. The superposed process $A \bigcup S$ is denoted A'.

Thinning captures inhibition effects. The probability of the deletion of A'_j, independently of all other flashes and perceived flashes, will be denoted by π_j. Two possibilities for π_j are considered: either **T.1**

$$\pi_j = D(A'_j - A'_{j-1})$$

or **T.2**

$$\pi_j = D(A'_{j+1} - A'_j)$$

where $D(.)$ is a nonincreasing function. In either case, this thinned process is denoted by A''.

They consider the case where the deletion probability $D(x)$ is the following simple piecewise constant function:

$$D(x) = \begin{cases} 1, & x < d \\ p, & x \geq d \end{cases} . \tag{2.5.1}$$

Here, $d \geq 0$, and $p \in [0, 1]$. If $d = 0$, deletions occur with probability p, completely independently of the flash process. If $d > 0$, some deletions occur independently of the process, but flashes or apparent flashes which are within d time units of their predecessor (alternatively, successor) are deleted with probability 1. In this case, there is nonlinear inhibition due to consecutive flashes occurring too close together in time.

Random translation is used to describe the formation of the response process from the flash process. That is, assuming the reaction time is distributed with density $f_V(v)$, each response is of the form $B_i = A_i'' + V_i$, where $\{V_i\}$ is a sequence of independent and identically distributed random variables with common density $f_V(v)$, mean μ, and variance σ^2. Some possibilities are the normal and lognormal. The latter has theoretical justification (Ulrich and Miller (1993)). However, it has been shown in Braun et al. (2003) that at parameter values estimated from the given data, the lognormal and normal densities are not statistically distinguishable. Hence, following Asimit and Braun (2005), we shall assume normality. In later chapters, we also assume a normal delay distribution in the parametric models that we develop for go-no go and choice RT experiments.

2.6 Nonparametric Intensity Estimation

Braun et al. (2003) use nonparametric estimation in conjunction with their parametric model for simple RT data to make inferences on the eye-brain-hand system. Similarities between nonparametric and parametric intensity estimates can indicate that common features in the nonparametric estimates are predicted by the parametric model. Furthermore, features distinct to the nonparametric estimates suggest characteristics that the model cannot predict. Thus, a parametric model can provide a

foundation for understanding the behavior of nonparametric intensity estimates.

Nonparametric intensity estimates are provided by Brillinger (1975a). In the case that a cumulant mixing condition is satisfied, Brillinger (1975a) shows that his estimates are asymptotically normal. Stationary point processes satisfying these mixing conditions are commonly referred to as "Brillinger mixing" or "B-mixing" (e.g. Heinrich (1988)).

An r vector-valued stationary point process is said to be *Brillinger mixing* if all cumulant densities of order $\nu \geq 2$ satisfy the condition that the integral of the absolute value of the cumulant density is finite (see Brillinger, 1975, Heinrich, 1988). Heinrich (1988) provides a sufficient condition for a stationary cluster point process to be Brillinger-mixing. We apply his result to prove that the RT point process data is Brillinger-mixing when our model assumptions are satisfied.

We now review Brillinger's estimates of intensities of orders one through three, as well as their asymptotic distributions. An estimate of the first-order response intensity is given by

$$\hat{p}_B = B([0,T])/T,$$

where T is the time of the last response ($T = \max(B_k)$).

Let $K(x)$ be a symmetric second order smoothing kernel with compact support, and h be an appropriately chosen bandwidth. Then the second-order nonparametric intensity estimates are given by

$$\hat{p}_{AB}(u) = \frac{1}{Th} \sum_i \sum_j K\left(\frac{u - (B_j - A_i)}{h}\right),$$

and

$$\hat{p}_{BB}(u) = \frac{1}{Th} \sum_i \sum_{j \neq i} K\left(\frac{u - (B_j - B_i)}{h}\right).$$

We have found that the biweight kernel $K(x) = (15/16)(1 - x^2)^2 1_{\{|x|<1\}}$ works well (see for example, Wand and Jones (1995)).

Under Brillinger's mixing conditions the intensity function estimates are asymptotically normal as $h \to 0$ and $Th \to \infty$, with mean

$$E[\hat{p}_{AB}(u)] = \int K(u_1) p_{AB}(u - u_1 h)\, du_1 \tag{2.6.1}$$

and asymptotic variance

$$(Th)^{-1}p_{AB}(u). \tag{2.6.2}$$

The bias is of order $O(h^2)$.

A nonparametric estimate of $p_{AAB}(u,v)$ is

$$\hat{p}_{AAB}(u,v) = T^{-1}\sum_i\sum_{j\neq i}\sum_k K_h((u,v)-(B_k-A_i,B_k-A_j)), \tag{2.6.3}$$

where $K_h(x,y) = h^{-2}K(x/h,y/h)$, is a bivariate kernel function with bandwidth h. We use the product form of the kernel

$$K(u,v) = K(u)K(v).$$

When the property of Brillinger-mixing holds this estimate is asymptotically normal with mean

$$E[\hat{p}_{AAB}(u,v)] = \int_{-1}^{1}\int_{-1}^{1} K(u_1)K(v_1)p_{AAB}(u-u_1h, v-v_1h)du_1dv_1$$

and approximate variance $p_{AAB}(u,v)/(Th^2)$.

2.6.1 Bandwidth Selection

The choice of bandwidth is critical in the estimation of intensities and densities due to the variance-bias trade-off (Wand and Jones (1995)). When the bandwidth is too small the result is an undersmoothed estimate, which has little bias, but high variability since few points are used in the averaging process at each point. Such an estimate takes into account only the particular data set, and not the variation across samples. The other extreme is too large of a bandwidth, which yields an oversmoothed estimate that removes the essential structure of the underlying intensity. An oversmoothed estimate has small variance and large bias.

Optimal bandwidths for the intensities are found by a popular approach of minimizing the asymptotic mean integrated squared error (AMISE) with respect to h, the bandwidth (see Wand and Jones (1995)). The AMISE is a large sample approximation to the mean integrated squared error (MISE), which is the integral of the mean

squared error (MSE). The MSE of an estimate can be shown to be the sum of the estimate's variance and square of the bias.

We now derive the general form of the AMISE optimal bandwidth for a second-order intensity. Using the expression for the asymptotic mean (2.6.1) with a second order Taylor expansion of $p_{AB}(u - u_1 h)$ about u, together with

$$\int K(x)dx = 1, \quad \int xK(x)dx = 0, \quad \int x^2 K(x)dx < \infty,$$

and the notation of Wand and Jones (1995):

$$\mu_2(K) = \int x^2 K(x)\, dx, \quad R(K) = (\int K^2(x)\, dx)^2,$$

we get

$$E[\hat{p}_{AB}(u)] = p_{AB}(u) + \frac{h^2}{2}\frac{\partial^2}{\partial^2 u} p_{AB}(u)\mu_2(K) + o(h^2). \tag{2.6.4}$$

We obtain an expression for $MSE(\hat{p}_{AB}(u))$ by using (2.6.4) to obtain the bias, and Brillinger's approximation of the variance (2.6.2), yielding:

$$\begin{aligned} & MSE(\hat{p}_{AB}(u)) \\ & \doteq \frac{h^4}{4}\left(\frac{\partial^2}{\partial^2 u} p_{AB}(u)\right)^2 \mu_2^2(K) + (Th)^{-1} R(K) \int p_{AB}(u)\, du + o(h^4). \end{aligned} \tag{2.6.5}$$

Integrating over u in (2.6.5) yields $MISE(\hat{p}_{AB})$ from which we get the following expression for the AMISE:

$$AMISE(\hat{p}_{AB}) = \frac{h^4}{4}\mu_2^2(K) \int \left(\frac{\partial^2}{\partial^2 u} p_{AB}(u)\right)^2 du + (Th)^{-1} R(K) \int p_{AB}(u)\, du.$$

Minimizing AMISE($\hat{p}_{AB}$) with respect to h yields the optimal bandwidth h_{AMISE} for p_{AB}, given as follows.

$$h_{AMISE}(p_{AB}) = \left[\frac{R(K) \int p_{AB}(u)\, du}{T\mu_2^2(K) \int (\frac{\partial^2}{\partial^2 u} p_{AB}(u))^2\, du}\right]^{1/5}.$$

The integrals involving the intensity can be numerically approximated using the Trapezoid quadrature formula. In the case of the biweight kernel we have $\mu_2(K) = 1/7$ and $R(K) = 5/7$, so that

$$h_{AMISE}(p_{AB}) = \left[\frac{35 \int p_{AB}(u)\, du}{T \int (\frac{\partial^2}{\partial^2 u} p_{AB}(u))^2\, du}\right]^{1/5}. \tag{2.6.6}$$

The above expression for the optimal bandwidth holds for each estimate, replacing p_{AB} by the desired second-order intensity.

2.7 Modelling with Latent Variables

A maximum likelihood approach to the identification of nonlinear systems is described in Brillinger (1988b) and Brillinger (1988a), where a threshold-type model is introduced. His work is concerned with modelling the firing of a neuron either spontaneously or under the influence of other neurons. The firing of a neuron is described as dependent on some internal potential formed from a combination of internal and external mechanisms. When the internal potential exceeds some threshold, the neuron fires. The coefficients of the internal potential relate to the kernels in the expansion of $\mu_B(t)$. Details of the model are provided in Section 3.3, where we modify the model and apply it to simulated and real simple RT data.

The threshold models that we consider can be viewed as latent variable models. A complete review of latent variable models is given by Skrondal and Rabe-Hesketh (2004). In our models we discretize the times of stimuli and responses over a grid of time intervals. The observed responses are the discretized response random variables Y_t, and the noisy internal potentials $U_t + \varepsilon_t$ are the latent responses. The Y_t are ordinal responses that depend on the value of the unobservable $U_t + \varepsilon_t$, which we denote by Y_t^*. Models for U_t are discussed for different RT experiments in Chapters 3, 4, and 6.

In these chapters we set up the observed response Y_t to be an indicator random variable for a response occurring in a time interval with right endpoint t. Thus, Y_t takes on values in $\{0, 1\}$. We model the relationship between observed and latent responses by

$$Y_t = \begin{cases} 1, & Y_t^* > \theta \\ 0, & \text{otherwise} \end{cases}, \tag{2.7.1}$$

where Y_t^* is a noisy internal state variable, and θ is a threshold. The models that we propose for the internal state variable involve the times of the flashes.

In the case that a normal error distribution, respectively logistic error distribution, is assumed for the error term ε_t, our model corresponds to a dichotomous probit model, respectively dichotomous logit model. The logistic cumulative distribution

function is

$$F(x) \equiv \Pr(\varepsilon_t < x) = \frac{e^x}{1+e^x}.$$

In order to clarify the connection between the latent response and generalized linear model formulations we consider the probability p_t of a response at t conditional on the history of the process. We denote the history of the process up to time t by H_t, and we have

$$p_t = \Pr(Y_t = 1|H_t) = \Pr(y_t^* > \theta|H_t) = \Pr(U_t + \varepsilon_t > \theta) = \Pr(\varepsilon_t < U_t - \theta),$$

where the last equality holds only in the case that the distribution of ε_t is symmetric about zero. Since Y_t is Bernoulli distributed with parameter p_t, the likelihood function of the data is given by

$$\prod_t p_t^{Y_t}(1-p_t)^{1-Y_t},$$

and the latent response model coincides with a generalized linear model of the binomial family with a link function coinciding with the canonical link for the distribution family. For example, the link function for the glm formulation of the dichotomous logit model is the logit link. The link function provides the link between the random and deterministic components of the response variable Y. It is a function of the mean of the response variable and is linearly related to the explanatory variables. Estimates of θ and the parameters that U_t is a function of can then be found using maximum likelihood.

In Chapter 6 we also consider the case in which the response variable takes on three values, and as before, the latent response has the form $Y_t^* = U_t + \varepsilon_t$. The observed response y_t takes on one of three response categories $\{-1, 0, 1\}$, where $y_t = -1$, respectively $y_t = 1$, indicates a black response, respectively white response, occurring in a time interval with right endpoint t. We model the relationship between the observed and latent responses by

$$y_t = \begin{cases} -1, & \text{if } U_t + \varepsilon_t < \theta_B \\ 0, & \text{if } \theta_B < U_t + \varepsilon_t < \theta_W \\ 1, & \text{if } U_t + \varepsilon_t > \theta_W \end{cases}.$$

This set-up corresponds to an ordinal logit regression model, in the case that the error distribution is logistic.

Chapter 3

Modelling of Simple RT Data

Simple RT provides an indication of the complexity of the operations of the brain which are involved in performing a visual-motor task. In a simple RT experiment, flashes are presented to a subject, and immediately upon perception of a flash the observer hits a button. The times of the flashes and responses are both recorded. The observer might press the button in the absence of a flash because of internal noise within the eye-brain-hand system that causes the observer to mistakenly perceive a stimulus.

A parametric model for simple RT was presented by Braun et al. (2003), and we continue to study that model, as well as to introduce a threshold model for simple RT. Both models are of interest, but for different reasons. A parametric model provides a basis for understanding the behavior of nonparametric intensity estimates, while the parameters in a threshold model have direct biological interpretations regarding inferences on the eye-brain-hand system.

Two types of simple RT data are studied in this chapter. In the first type there is a single flash type, and the observer's task is to hit a button immediately upon perception of a flash. The simple RT data consists of the times of the flashes and the responses. Such data was used by Braun et al. (2003) and Asimit and Braun (2005). An alternative way to run a simple RT experiment is to present two flash types (black and white), and the subject hits a button immediately after perception of either flash type; the same response occurs regardless of the flash type. For this case, in addition

to the times of the flashes and the responses, the simple RT data also consists of the flash type at each time.

Braun et al. (2003) use their parametric model to develop hypothesis tests involving second-order intensities. Using the same model, Asimit and Braun (2005) study the third order intensity $p_{AAB}(u, v)$, and propose hypothesis testing methodology based on features of $p_{AAB}(u, v)$ observed in contour plots. The parametric model and testing methodology can also be applied to the simple RT data with two flash types. If the subject has a tendency to respond to one flash type more than the other, then we say that the subject has a flash-type bias. Due to the fact that the same response is made to both flash types, we make the simplifying assumption that there is no flash-type bias. This assumption allows us to consider only the times of the flashes, which in turn, enables us to apply the parametric model to both types of simple RT data that we consider in this chapter.

In Section 2.5 we reviewed the parametric model of Braun et al. (2003), and in Section 3.0.1 we review tests from Braun et al. (2003) and Asimit and Braun (2005). In Section 3.1, we give the details of the covariance approximation of a third order intensity at two points, which is used in the hypothesis testing of Asimit and Braun (2005). The parameter estimation of Braun et al. (2003) is reviewed and modified slightly in Section 3.2, where we compare it with nonlinear regression as an alternative method of estimation when it is of interest to obtain a single set of parameter estimates for data sets obtained using different flash rates. A threshold model for simple RT is introduced in Section 3.3, which includes diagnostics and applications to various simulated and real data sets. A random effects threshold model is proposed in Section 3.4, as an alternative to fitting a threshold model to pooled data. Finally, a discussion of the methods and models considered in this chapter follows in Section 3.5.

3.0.1 Testing of Thinning Assumptions

We denote the number of A events, respectively B events, in the interval X by $A(X)$, respectively $B(X)$. The results of Brillinger (1975a), provided in Section 2.6, are

used to obtain nonparametric intensity estimates. Such estimates are asymptotically normal under mixing assumptions given by Brillinger (1975a), which we will show are satisfied for the simple RT model, in Section 3.1.

Plots of the second-order intensities $p_{BB}(u)$ and $p_{AB}(u)$ are studied by Braun et al. (2003), and based on features of these plots they develop their hypothesis tests. The function $p_{BB}(u)$ is the same under thinning assumptions T.1 and T.2, and has a trough centered at 0, which widens and deepens as d increases; as d approaches zero the curve flattens out, and is constant when $d = 0$ (no nonlinear inhibition). Based on these observations, they use pointwise bootstrapped confidence bands for $E[\hat{p}_{BB}(u)]$ to test for nonlinear inhibition. In the absence of nonlinear inhibition, $p_{BB}(u)$ and $E[\hat{p}_{BB,h}(u)]$ should be constant; thus, if a horizontal line cannot be contained within the confidence bands, then there is evidence of nonlinear inhibition.

The effect of the d parameter on $p_{AB}(u)$ is to depress the region to the right of the peak when thinning assumption T.1 is operative, and to depress the region to the left of the peak when T.2 is operative. Thus, Braun et al. (2003) use confidence bands for $E[\hat{p}_{AB,h}(u)]$ to test the appropriateness of each thinning assumption. If the null hypothesis is T.1, then one obtains confidence bands for $E[\hat{p}_{AB,h}(u)]$ and overlays a graph of $E[\hat{p}_{AB,h}(u)]$ computed under the T.1 assumption and evaluated at the parameter estimates. If the overlaid curve falls outside the confidence bands, there is evidence against the assumption T.1 as postulated by the model. In a similar manner, the null hypothesis of T.2 can be tested.

The tests of Braun et al. (2003) have a high degree of dependency on the model. If the hypothesis T.1 is rejected, the conclusion is not that there is no thinning of that type. The only conclusion that can be made is that there is no thinning of that type, under the assumption that the model is otherwise correct. Another deficiency of that test is that it has a very low power for $d < .25$ (when flashes are presented at rate ($p_A = 1$)). Thus, Asimit and Braun (2005) sought more powerful tests which do not have the same degree of dependence on the model. These tests are based on the third order intensity $p_{AAB}(u, v)$, and they consider two tests of hypotheses: T1 versus T.2 and T.2 versus T.1. Asimit and Braun (2005) derive the following expressions

for $p_{AAB}(u,v)$:

$$p_{AAB}(u,v) = (1-p)p_A^2 e^{-dp_A}[p_A q_- + g_+(v,u,d) + g_+(u,v,d)] \tag{3.0.1}$$

under thinning assumption **T.1**, and

$$p_{AAB}(u,v) = (1-p)p_A^2 e^{-dp_A}[p_A q_+ + g_-(v,u,d) + g_-(u,v,d)] \tag{3.0.2}$$

under assumption **T.2**, where

$$\begin{aligned} q_- &= P(V \notin (u-d,u), V \notin (v-d,v)), \\ q_+ &= P(V \notin (u,u+d), V \notin (v,v+d)), \\ g_-(v,u,d) &= I(v \notin (u-d,u))f(u), \end{aligned}$$

and

$$g_+(v,u,d) = I(v \notin (u,u+d))f(u).$$

Motivation for their test statistic is based on characteristics observed in contour plots of $p_{AAB}(u,v)$ under each thinning assumption. Under T.1, a trough appears near the upper right side of the peak (near (μ,μ)). An identical trough appears under T.2, but near the lower left side of the peak.

The test statistic is formed by considering the trough under the thinning assumption of the alternative hypothesis, and choosing two points. One point (u_L, v_L) falls within the trough and near the peak, while (u_U, v_U) is directly above (u_L, v_L) (i.e. $u_L = u_U$) in the direction away from the peak. Under the null hypothesis, we have $p_{AAB}(u_U, v_U) \approx p_{AAB}(u_L, v_L)$, and $p_{AAB}(u_U, v_U) > p_{AAB}(u_L, v_L)$ under the alternative hypothesis, which leads to a one-sided test. The test points are chosen so that the power function is maximized for one data set, and for subsequent data sets the test points are shifted by the difference in the RT mean estimates $\hat{\mu}$, in order to avoid bias. The test statistic of interest is:

$$\frac{\hat{p}_{AAB}(u_U, v_U) - \hat{p}_{AAB}(u_L, v_L)}{s_0}, \tag{3.0.3}$$

where $\hat{p}_{AAB}(u,v)$ is a nonparametric intensity estimate of $p_{AAB}(u,v)$, and s_0 is a consistent estimator of the standard error of $\hat{p}_{AAB}(u_U, v_U) - \hat{p}_{AAB}(u_L, v_L)$ when the

null hypothesis is true. The variance estimate s_0^2 is a linear combination of two variances and a covariance. The variances and covariance are approximated, but with modifications to reduce variability, bias, and computation time. Some details of the modifications follow.

Based on simulations, Prokop (2004) found that Brillinger's variance approximation, as given in Section 2.6, underestimates the variance, and at certain test points there is a high variability in the approximations. This was overcome by modifying Brillinger's estimator so that $p_{AAB}(u, v)$ is replaced by $E[\hat{p}_{AAB}(u, v)]$. Brillinger's approximation for the expectation is provided in Section 2.6. Asimit and Braun (2005) estimate expectation approximations using the Trapezoid Quadrature Rule. An approximation of the covariance will be derived in Section 3.1. Comparisons of $V(\hat{p}_{AAB}(u_L, v_L) - \hat{p}_{AAB}(u_U, v_U))$ based on simulations and the expression using the covariance approximation indicated that the approximation underestimates the variance of the difference by an approximate factor of 1.3 (see Prokop (2004)). In order to correct the variance estimation of $\hat{p}_{AAB}(u_L, v_L) - \hat{p}_{AAB}(u_U, v_U)$ (evaluated under the null thinning assumption) Asimit and Braun (2005) use 1.3 as a correction factor for the variance.

For moderate values of the thinning parameter d, the test is found to have moderately high power, but the results are inconclusive. For each of the twelve data sets, Asimit and Braun (2005) fail to reject T.2 in favor of T.1. In the test of T.1 vs. T.2, two of the twelve p-values are moderately small, but such results are consistent with T.1 for each data set. Due to the inconclusiveness of their results, neither T.1 nor T.2 can be ruled out for any of the data sets, and we cannot rule out the possibility that a thinning mechanism different from T.1 and T.2 may be in effect.

3.1 Approximation of $\mathbf{Cov}(\hat{p}_{AAB}(u_1, v_1), \hat{p}_{AAB}(u_2, v_2))$ for a Simple RT Parametric Model

The test statistic (3.0.1) of Asimit and Braun (2005) is evaluated at points chosen such that (u_L, v_L) falls within the trough and near the peak, while (u_U, v_U) is directly above (u_L, v_L) (i.e. $u_L = u_U$) in the direction away from the peak. In this section we determine an approximation for the covariance between $\hat{p}_{AAB}(u_1, v_1)$ and $\hat{p}_{AAB}(u_2, v_2)$, when either $u_1 = u_2$ or $v_1 = v_2$, and the points satisfy

$$\max\{u_1 - v_1, u_1 - v_2\} \geq 2h \text{ or } \max\{v_1 - u_1, v_1 - u_2\} \geq 2h, \tag{3.1.1}$$

respectively. This approximation is valid under mixing conditions given by Brillinger (1975a), which we show are satisfied.

3.1.1 Simple RT as a Brillinger-mixing Process

Heinrich (1988) provides a sufficient condition for a stationary cluster point process to be Brillinger-mixing. A cluster point process consists of a point process of cluster centers (primary process), and a point process of cluster members (secondary process). Each cluster center generates a point process of cluster members, and the superposition of the cluster members forms the cluster point process. Let $N_s(X)$ denote the number of secondary process points in the set X. Heinrich's sufficient condition is that the point process of cluster centers is Brillinger-mixing, and $E[N_s^k(X)] < \infty$, for all bounded subsets X, and $\forall k \geq 1$. In the case of a Poisson cluster process, the cluster centers are a Poisson process, which is Brillinger-mixing. Thus, it only remains to show that $E[N_s^k(X)] < \infty$, for all bounded subsets X, and $\forall k \geq 1$.

For our simple RT model with no superposition ($p_S = 0$), the response process B can be viewed as a Poisson cluster process with the flash process A as the primary process (see Section 2.4). We denote A_i to be the ith flash, which is the ith cluster center. Each cluster center A_i generates either $\{\emptyset\}$ or $\{A_i + V_i\}$, where the V_i are identically distributed with density $f_V(\cdot)$. Therefore, $E[N_s^k(X)] \leq 1$. Thus, the

response process is Brillinger-mixing. In the case that $p_S > 0$ the same argument holds with the A process replaced by the Poisson process A' with rate $p_A + p_S$.

In order to show that the bivariate point process (A, B) is Brillinger-mixing, we consider the point process $A \bigcup B$, which we denote by C. The process C is a Poisson cluster process with cluster centers at A_i. When there is no thinning in our model (i.e. $d = 0$ and $p = 0$), each cluster center A_i generates two points: $\{A_i + 0, A_i + V_i\}$, where the V_i are as before. In the case that there is thinning, there is also the possibility that A_i generates only one point: $\{A_i + 0\}$. Therefore, regardless of the presence or absence of thinning, $N_s(X)$ takes on values in $\{1, 2\}$, and it follows that $E[N_s^k(X)] \leq 2^k < \infty$. Thus, $A \bigcup B$ is Brillinger-mixing, from which we conclude that the bivariate point process (A, B) is also Brillinger-mixing.

3.1.2 Derivation of a $\mathbf{Cov}(\hat{p}_{AAB}(u_1, v_1), \hat{p}_{AAB}(u_2, v_2))$ Approximation

Before we proceed with the derivation of the covariance approximation, we state our result. The covariance approximation under constraint (3.1.1) is given by

$$\begin{aligned} &Cov(\hat{p}_{AAB}(u_1, v_1), \hat{p}_{AAB}(u_2, v_2)) \qquad (3.1.2)\\ &\doteq \frac{1}{Th^2} \int\int k_2(x, y, u_1, u_2, v_1, v_2, h)\, dx\, dy \\ &+\frac{1}{Th} \int\int\int (k_3(x, y, z, u_1, u_2, v_1, v_2, h) + k_3(y, x, z, v_1, v_2, u_1, u_2, h))\, dx\, dy\, dz, \end{aligned}$$

where

$$\begin{aligned} &k_2(x, y, u_1, u_2, v_1, v_2, h) \\ &\quad = K(x)K(y)K\left(x + \frac{u_2 - u_1}{h}\right)K\left(y + \frac{v_2 - v_1}{h}\right)p_{AAB}(u_1 - xh, v_1 - yh), \qquad (3.1.3) \end{aligned}$$

and

$$\begin{aligned} &k_3(x, y, z, u_1, u_2, v_1, v_2, h) \\ &\quad = K(x)K(y)K\left(x + \frac{u_2 - u_1}{h}\right)K(z)p_{AAAB}(u_1 - xh, v_1 - yh, v_2 - zh). \qquad (3.1.4) \end{aligned}$$

Our approximation can be simplified further by using a Taylor expansion of $p_{AAB}(u, v)$, but this was found to lead to numerical problems such as asymmetry:

$\hat{p}_{AAB}(u,v) \neq \hat{p}_{AAB}(v,u)$. In the same manner as the derivation of $p_{AAB}(u,v)$ (see Asimit and Braun (2005)), expressions for $p_{AAAB}(t,u,v)$ can be found. For our parametric model,

$$p_{AAAB}(t,u,v) = (1-p)p_A^3 e^{-dp_A}[p_A r_- + s_+(t,u,v,d) + s_+(u,v,t,d) + s_+(t,v,u,d)]$$

under thinning assumption T.1, and

$$p_{AAAB}(t,u,v) = (1-p)p_A^3 e^{-dp_A}[p_A r_+ + s_-(t,u,v,d) + s_-(u,v,t,d) + s_-(t,v,u,d)]$$

under assumption T.2, where

$$r_- = P(V \notin (t-d,t), V \notin (u-d,u), V \notin (v-d,v)),$$
$$r_+ = P(V \notin (t,t+d), V \notin (u,u+d), V \notin (v,v+d)),$$
$$s_-(t,u,v,d) = I(t \notin (v-d,v), u \notin (v-d,v))f(v),$$

and

$$s_+(t,u,v,d) = I(t \notin (v,v+d), u \notin (v,v+d))f(v).$$

As in the variance approximation, $p_{AAB}(u,v)$ is replaced by $\mathrm{E}[\hat{p}_{AAB}(u,v)]$ in the integral involving $k_2(x,y,u_1,u_2,v_1,v_2,h)$ in the covariance approximation (3.1.2). We numerically approximate this integral by implementing the Trapezoid Quadrature formula. Similarly, $p_{AAAB}(t,u,v)$ can be replaced by $\mathrm{E}[\hat{p}_{AAAB}(t,u,v)]$ in integrals involving $k_3(x,y,z,u_1,u_2,v_1,v_2,h)$. However, it is too time-consuming to obtain $\mathrm{E}[\hat{p}_{AAAB}(t,u,v)]$. Therefore, in order to increase the computation speed, we use Monte Carlo integration to approximate the integrals involving $k_3(x,y,z,u_1,u_2,v_1,v_2,h)$.

We now derive an approximation of the covariance. The covariance is given by

$$\begin{aligned}
&Cov(\hat{p}_{AAB}(u_1,v_1), \hat{p}_{AAB}(u_2,v_2)) \\
&= E[\hat{p}_{AAB}(u_1,v_1)\hat{p}_{AAB}(u_2,v_2)] - E[\hat{p}_{AAB}(u_1,v_1)]E[\hat{p}_{AAB}(u_2,v_2)] \\
&= \frac{1}{T^2h^4}\int_{[0,T]^6}\cdots\int \kappa_4(u_1,v_1,u_2,v_2,x_1,y_1,z_1,x_2,y_2,z_2) \\
&\quad\times(E[A(dx_1)A(dx_2)A(dy_1)A(dy_2)B(dz_1)B(dz_2)] \\
&\quad -E[A(dx_1)A(dy_1)B(dz_1)]E[A(dx_2)A(dy_2)B(dz_2)]),
\end{aligned}$$

where

$$\kappa_4(u_1, v_1, u_2, v_2, x_1, y_1, z_1, x_2, y_2, z_2)$$
$$= K\left(\frac{u_1 - z_1 + y_1}{h}\right) K\left(\frac{v_1 - z_1 + x_1}{h}\right) K\left(\frac{u_2 - z_2 + y_2}{h}\right) K\left(\frac{v_2 - z_2 + x_2}{h}\right). \quad (3.1.5)$$

Under constraint (3.1.1), many of the cross-product terms of $E[\hat{p}_{AAB}(u_1, v_1)\hat{p}_{AAB}(u_2, v_2)]$ are zero due to the biweight kernel having support in $[-1, 1]$. For example, when $x_1 = y_1$, $x_2 = y_2$, and $z_1 = z_2$, the product of the kernels (3.1.5) is 0. Since either $u_1 = u_2$ or $v_1 = v_2$, and $p_{AAB}(u, v)$ is symmetric, without loss of generality we assume that $u_1 = u_2$. Thus, (3.1.5) becomes

$$K\left(\frac{u_1 - z_1 + y_1}{h}\right) K\left(\frac{v_1 - z_1 + y_1}{h}\right) K\left(\frac{u_1 - z_1 + y_2}{h}\right) K\left(\frac{v_2 - z_1 + y_2}{h}\right),$$

and in the case that $u_1 - v_1 \geq 2h$, it follows that $\frac{u_1 - z_1 + y_1}{h} \geq \frac{v_1 - z_1 + y_1}{h} + 2$. Thus, $K\left(\frac{u_1 - z_1 + y_1}{h}\right)$ and $K\left(\frac{v_1 - z_1 + y_1}{h}\right)$ cannot be simultaneously nonzero. A similar argument can be used in the case that $u_1 - v_2 \geq 2h$. Because of asymptotic independence the differences between the remaining terms with terms from $E[\hat{p}_{AAB}(u_1, v_1)]E[\hat{p}_{AAB}(u_2, v_2)]$ are negligible.

Since the bivariate point process (A, B) is Brillinger mixing, we conclude that all intensities of (A, B) are bounded. The basic approximation involves terms of order $(Th^2)^{-1}$ (the p_{AAB} term), and other terms can be included to improve the approximation; since the p_{AAAB} terms (have order $(Th)^{-1}$) are simple to work with, we include them. The resulting approximation is

$$\frac{1}{T^2h^4}\{\int_0^T \int_0^T \int_0^T \kappa_4(u_1, v_1, u_2, v_2, x_1, y_1, z_1, x_1, y_1, z_1) E[A(dy_1)A(dx_1)B(dz_1)] +$$
$$\int_0^T \int_0^T \int_0^T \int_0^T \kappa_4(u_1, v_1, u_2, v_2, x_1, y_1, z_1, x_2, y_1, z_1) E[A(dy_1)A(dx_1)A(dx_2)B(dz_1)] +$$
$$\int_0^T \int_0^T \int_0^T \int_0^T \kappa_4(u_1, v_1, u_2, v_2, x_1, y_1, z_1, x_1, y_2, z_1) E[A(dy_1)A(dy_2)A(dx_1)B(dz_1)]\}.$$

By making the change of variables $t_1 = z_1 - y_1, t_2 = z_1 - x_1, t_3 = z_1 - x_2, t_4 = z_1$ in the first integral we obtain

$$\frac{1}{T^2h^4} \int_0^T \int_{-T}^T \int_{-T}^T K\left(\frac{u_1 - t_1}{h}\right) K\left(\frac{v_1 - t_2}{h}\right) K\left(\frac{u_2 - t_1}{h}\right) K\left(\frac{v_2 - t_2}{h}\right)$$
$$\times E[A(d(t_3 - t_1))A(d(t_3 - t_2))B(dt_3)],$$

which, by the definition of $p_{AAB}(u,v)$, simplifies to

$$\frac{1}{Th^4}\int_{-T}^{T}\int_{-T}^{T} K\left(\frac{u_1-t_1}{h}\right)K\left(\frac{v_1-t_2}{h}\right)K\left(\frac{u_2-t_1}{h}\right)K\left(\frac{v_2-t_2}{h}\right)p_{AAB}(t_1,t_2)dt_1dt_2,$$

and another change of variables, $x=\frac{u_1-t_1}{h}, y=\frac{v_1-t_2}{h}$, yields

$$\frac{1}{Th^2}\int_{\frac{u_1-T}{h}}^{\frac{u_1+T}{h}}\int_{\frac{v_1-T}{h}}^{\frac{v_1+T}{h}} K(x)K(y)K\left(x+\frac{u_2-u_1}{h}\right)K\left(y+\frac{v_2-v_1}{h}\right)p_{AAB}(u_1-xh,v_1-yh)dxdy.$$

The second integral in the covariance approximation is simplified in a similar manner, with the change of variables $t_1=z_1-y_1, t_2=z_1-x_1, t_3=z_1-x_2, t_4=z_1$ followed by the change of variables $x=(u_1-t_1)/h$, $y=(v_1-t_2)/h$, $z=(v_2-t_3)/h$. Therefore, under constraint (3.1.1) an approximation of the covariance between two third order intensity estimates is given by

$$Cov(\hat{p}_{AAB}(u_1,v_1),\hat{p}_{AAB}(u_2,v_2)) \doteq \frac{1}{Th^2}\int\int k_2(x,y,u_1,u_2,v_1,v_2,h)\,dx\,dy$$
$$+\frac{1}{Th}\int\int\int (k_3(x,y,z,u_1,u_2,v_1,v_2,h)+k_3(y,x,z,v_1,v_2,u_1,u_2,h))\,dx\,dy\,dz,$$

where $k_2(\cdot)$ and $k_3(\cdot)$ are given respectively by (3.1.3) and (3.1.4).

3.2 Parameter Estimation for a Simple RT Parametric Model

In this section we discuss and compare two methods of parameter estimation for simple RT data, using real and simulated data. First, we consider a modification of the method of Braun et al. (2003); we refer to this modified method as the original method. We also introduce an estimation method that is based on nonlinear regression. This proposed method can be used when data is available from experiments run at different flash rates, and it is believed that the parameters have the same values across the data sets. We implement a simulation study to compare the performance of the two estimation methods in terms of their bias, variance and MSE.

We apply our methods to simple RT data in which black and white flashes are presented, and the observer presses a button immediately upon perception of either

flash type. At each of nine different flash rates, a simple RT experiment was run eight times, where each run consisted of 100 flashes. For the parameter estimation we pool the runs of the same rate into one data set, and consider estimation for the nine pooled data sets. Pooling is done by concatenating consecutive runs together so that the last event of run j is followed by the first event of run $j+1$, with the times adjusted appropriately.

3.2.1 Original Estimation Method

In Braun et al. (2003) an iterative algorithm is used to obtain $\hat{\mu}$, $\hat{\sigma}$, and $\hat{p}_S$ for simple RT data. In their algorithm, estimates of μ and σ are obtained after each iteration. However, when the delay distribution is symmetric, the location of the peak of the intensity $p_{AB}(u)$ coincides with the mean of the reaction times. Thus, in such cases, the location of the peak for the nonparametric estimate of $p_{AB}(u)$ should be near the value of the mean reaction time. Therefore, since we are assuming a $N(\mu, \sigma)$ delay distribution, we fix the peak location as an initial value for the estimation of μ. The estimates of σ and the noise rate p_S are as in their algorithm. The modified algorithm is given as Algorithm 3.2.1.

Algorithm 3.2.1. *For RT data (A, B) the parameters μ, σ, and p_S can be estimated as follows:*

1. *Find the location of the peak in a nonparametric intensity estimate of $p_{AB}(u)$ and set μ_0 as the peak location.*

2. *Match each response time B_j with a flash time preceding it, $\hat{A}''_j$, such that $B_j - \hat{A}''_j$ is closest to μ_0, to obtain estimates of the reaction times: $\hat{V}_j = B_j - \hat{A}''_j$.*

3. *Estimate μ and σ by the mean and standard deviation of the $\hat{V}_j$.*

4. *Let $T = B_{N_B}$, where N_B is the number of responses. Estimate the noise rate p_S by:*

$$\hat{p}_S = \frac{|\{\hat{A}''_j : \hat{A}''_j = \hat{A}''_k\}|}{T},$$

where $|\cdot|$ denotes set cardinality.

Braun et al. (2003) showed that for the simple RT parametric model the response rate can be expressed as

$$p_B = (1-p)(p_A + p_S)e^{-d(p_A+p_S)}. \tag{3.2.1}$$

They find estimates of d and p by manipulating (3.2.1) to obtain an expression for p that involves d:

$$\hat{p} = 1 - \frac{\hat{p}_B}{(p_A + \hat{p}_S)e^{-d(p_A+\hat{p}_S)}}, \tag{3.2.2}$$

where $\hat{p_S}$ is replaced by the estimate obtained from Algorithm 3.2.1, and $\hat{p_B} = N_B(T)/T$ is the nonparametric estimate of the response rate (see Section 2.6). It then follows that the second order intensity $p_{BB}(u) = p_B^2 P(|V_2 - V_1 - u| > d)$ can be expressed as a univariate function of d by replacing p with (3.2.2).

It was shown in Braun et al. (2003) that $p_{BB}(u)$ is useful in obtaining an estimate of d. This intensity has a trough at $u = 0$ for $d > 0$ and as d approaches 0 the trough flattens out, while as d increases the trough deepens. Furthermore, for long-run experiments (large T values) the bias of the nonparametric intensity estimate is near zero. Thus, the estimation of d is based on minimizing the difference between the parametric (based on model expression with parameter estimates) and corresponding nonparametric estimate. By taking advantage of the fact that this intensity has a trough at lag $u = 0$ an estimate of d is obtained by finding $\hat{d}$ that minimizes the distance of the two intensity estimates evaluated at lag 0.

3.2.2 Nonlinear Regression Estimation

The dependent variable y in a nonlinear regression method is modeled by a nonlinear function of an independent variable x with parameters $\underline{\theta}$ (see Montgomery et al. (2001)). Suppose the simple RT model parameters d, p, and p_S are the same for $k > 3$ data sets with different flash rates p_A. When this is the case, it is desirable to obtain a single set of parameter estimates $\{\hat{d}, \hat{p}, \hat{p_S}\}$ for the k data sets. Using the nonparametric response rate estimates $\hat{p}_B$ for each data set along with the set flash rates p_A, we may view (3.2.1) as a nonlinear regression model where d, p, and p_S are

unknown model parameters. Doing so takes into account the different response rates for the various flash rates and allows us to obtain simultaneous estimates for d, p, and p_S.

If it is the case that the nonlinear regression parameter estimates are not within the required ranges, $p_S > 0$, $0 < p < 1$, and $d > 0$, rather than getting direct parameter estimates, we work with (3.2.1) in terms of q, r, and s, where $q = \log(p_S)$, $r = \log(\frac{p}{1-p})$, and $s = \log(d)$. That is, the nonlinear regression model becomes

$$p_B = \frac{1}{1+e^r}(p_A + e^q)\exp(-e^s(p_A + e^q)),$$

and we estimate q, r, and s. We obtain estimates of p_S, p, and d by the inverse relations, so that $\hat{p}_S = e^{\hat{q}}, \hat{p} = \frac{e^{\hat{r}}}{1+e^{\hat{r}}}$, and $\hat{d} = e^{\hat{s}}$.

Under the assumption of equal parameters across data sets, in addition to obtaining a single set of parameter estimates, an advantage of using nonlinear regression to obtain parameter estimates is that it is simple to perform hypothesis tests on the parameters.

3.2.3 Performance Study of the Two Estimation Methods

In order to study the performance of the two estimation methods, in terms of bias, variance and mean squared error (MSE), we conduct a simulation study. Each simulation run consists of 17 simulations, where each simulation is comprised of N_A flashes with a different rate from $\{.4, .5, \ldots, 1.9, 2\}$, and with fixed parameters $d = .2$, $p = .15$, $p_S = 0$, $\mu = .4$, and $\sigma = .1$. One set of nonlinear regression estimates is obtained for each simulation run by the nature of the estimation method. Our nonlinear regression model is fit using the `nls` function in the statistical package R (R Development Core Team (2006)). For the original estimation method, a set of parameters is obtained for each simulation within a run. Therefore, since we are assuming that the parameters to be estimated are the same (which is a correct assumption in this case) the estimates within each run are averaged in order to be comparable with the single set of estimates from nonlinear regression.

As expected, when the number of flashes increases from 800 to 1200 there is a

Table 3.1: Performance comparison of the two estimation methods based on 1000 simulation runs of RT data sets with N_A flashes and parameters $d = .2$, $p = .15$, $p_S = 0$, $\mu = .4$, and $\sigma = .1$, and flash rates of $\{.4, .5, \ldots, 1.9, 2\}$.

		Bias		Variance		MSE	
	N_A	800	1200	800	1200	800	1200
original	d	-.009533	-.008887	.000027	.000015	.000112	.000094
	p	.017500	.016936	.000075	.000049	.000382	.000336
	p_S	.021120	.020451	.000002	.000001	.000448	.000420
regression	d	-.015611	-.010953	.001377	.000924	.001620	.001044
	p	.026030	.018230	.003316	.002119	.003994	.002451
	p_S	.022708	.016428	.001119	.000668	.001634	.000938

reduction in the bias, variance, and MSE of each estimator, regardless of the estimation method. Overall, the estimators resulting from the original method appear to be slightly better than those obtained from nonlinear regression; estimators from the original method appear to possess smaller bias, variance, and MSE than the corresponding estimates resulting from nonlinear regression. However, the differences in these performance measures are quite small; differences in bias, variance and MSE between the two methods are less than 10^{-2}.

Due to the similarities in bias, variance, and MSE for the two estimation methods, when the parameter values are assumed to be the same across data sets of different rates, we recommend nonlinear regression estimation because of its advantages. With nonlinear regression, as well as easily obtaining a single set of parameter estimates, hypothesis testing for the parameters is readily available. Thus, we conclude that an important experimental design consideration is to include at least four different flash rates (since estimating three parameters) when running simple RT experiments so that nonlinear regression may be used to obtain estimates of d, p, and p_S.

3.2.4 Application to Simple RT Data

We first find parameter estimates for the nine pooled simple RT data sets using the original estimation method. In order to avoid negative estimates of the thinning probability p, we set the estimate as $\max\{\hat{p}, 0\}$. The results are given in Table 3.2.

Table 3.2: Parameter estimates for each of the nine pooled data sets

flash rate, p_A	$\hat{\mu}$	$\hat{\sigma}$	$\hat{d}$	$\hat{p}$	$\hat{p_S}$
0.4	.296	.049	.065	.026	.020
0.6	.291	.049	.101	0	.019
0.8	.286	.053	.109	0	.035
1.0	.283	.057	.117	.007	.029
1.2	.282	.058	.110	.001	.063
1.4	.284	.066	.125	0	.058
2.0	.278	.064	.114	0	.122
4.0	.283	.070	.120	0	.134
8.0	.285	.055	.100	0	.181

Across the nine data sets, the means and variances of $\hat{d}$, $\hat{p}$, and $\hat{p}_S$ are (.107, .000313), (.003778, .000075), and (.073444, .003402), respectively, suggesting that these parameters are similar for the nine data sets. For this reason, we proceed with obtaining estimates of d, p, and p_S via nonlinear regression.

We obtain nonlinear regression estimates for the reaction time data in R using the function `nls`. The estimates are $\hat{d} = 0.093$, $\hat{p} = 0.041$, and $\hat{p}_S = 0.072$, with p-values of 3.1×10^{-6}, 0.341, and 0.193, respectively. This provides us with additional evidence that the thinning parameter d is nonzero, so that there is nonlinear inhibition present in the eye-brain-hand system. It also suggests that the thinning probability p is zero, which means that according to our model flashes are rarely deleted with complete randomness. Finally, it indicates that any internal noise process has a negligible rate, so that there is essentially no noise in the system.

There are only nine pairs (p_A, p_B) used in the regression, but as a quick adequacy check for the fit of the model a plot of the fitted response rates against the observed values (nonparametric estimates) is provided in Figure 3.1. It can be seen that the model is able to predict the response rates quite well.

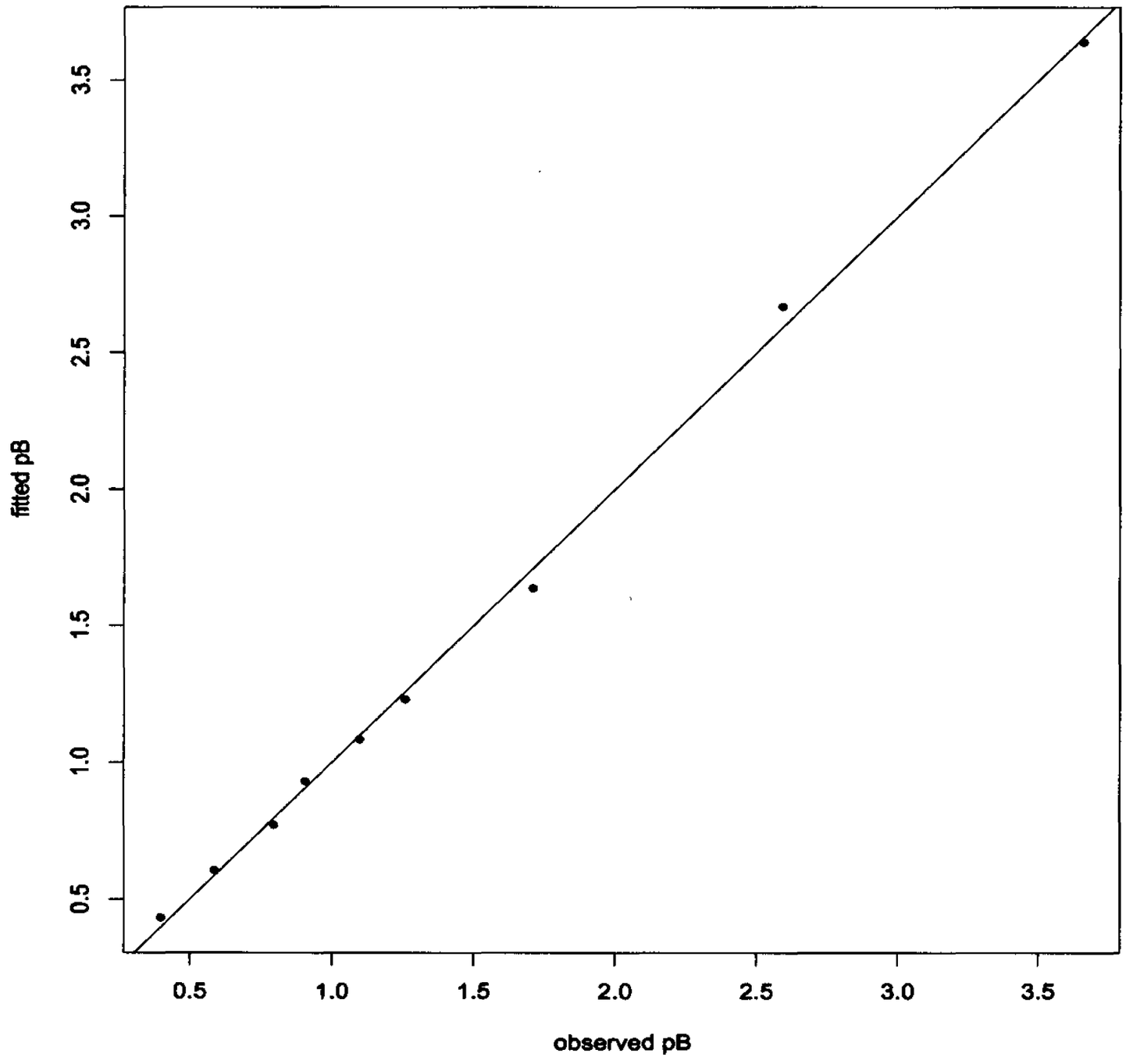


Figure 3.1: Plot of the fitted vs. observed response rates for the nonlinear regression model fit to nine pooled simple RT data sets.

3.3 Threshold Models for Simple RT

In the analysis of a simple RT experiment, another conceptual model of interest is a threshold model (integrate-and-fire model), (see Brillinger (1988b), Brillinger (1988a)). In such a model, a response occurs when an internal state variable exceeds a threshold. Brillinger used this type of model to analyze a spontaneously firing

neuron ("noise" input), as well as the firing of a neuron under the influence of two other neurons (Brillinger (1988b)). Estimates of the threshold and other internal quantities are obtained via maximum likelihood. Advantages of this approach include biologically interpretable parameters and the availability of formal hypothesis testing.

We first consider the case where only one flash type is presented. Let $A = \{A_1, A_2, \ldots\}$ represent the sequence of flash times, and $A(t)$ denote the number of flashes that occur in the time interval $(0, t]$. We denote the times of the responses by $B = \{B_1, B_2, \ldots\}$, with $B(t)$ defined analogous to $A(t)$. As in Brillinger (1988a), we discretize into a 0-1 valued time series for analysis. We transform the data by choosing a lag value m and creating time intervals $\{(0, m], (m, 2m], \ldots, (T, T+m]\} = \{I_1, I_2, \ldots, I_T\}$, where $T = m\lceil \max(B)/m \rceil$. The lag m may be chosen such that there is at most one flash (response) in each discretization interval, but to decrease computation time we choose m slightly larger, allowing more than one point in a given interval. We discretize the A process as follows:

$$X_j = \begin{cases} 1, & \text{if there exists } A_k \text{ such that } A_k \in I_j \\ 0, & \text{otherwise} \end{cases}.$$

We discretize the B process in a similar way:

$$Y_j = \begin{cases} 1, & \text{if there exists } B_k \text{ such that } B_k \in I_j \\ 0, & \text{otherwise} \end{cases}.$$

That is, we set X_j and Y_j to equal 1 or 0, depending on whether or not there is a flash or response, respectively, in interval I_j.

Following the terminology of Brillinger (1988b), a_u is a summation function that is defined as follows. Assuming that a flash occurs at time τ, $a_{t-\tau}$ represents the effect of that flash on the internal potential at time t. Both a_t and the first-order kernel $a_1(u)$, discussed in Section 2.3.2, estimate the average impulse response of the system.

Suppose $t \in I_j$, so that j is the index of the interval containing time t. As of time t, we denote the number of time intervals elapsed since the last interval containing a response by γ_j. Under the assumption $Y_0 = 1$, it follows that $\gamma_1 = 1$, and letting

k be the index of the interval in which the first response occurs, we have $\gamma_j = j$ for $j = 1, \ldots, k$. In general, $\gamma_j = j - i$ for $j = 1, \ldots, I_T$, where i is the index of the last interval before I_j that contains a response. Note that time is in terms of an index j corresponding to the interval I_j containing the time of interest t. An example of how we obtain γ_j for the first five time intervals when the first two responses occur at times 0.12 s and 0.23 s, and the time lag is $m = .05$ is given in Table 3.3.

Table 3.3: Example of γ_t calculations

j	Y_j	γ_j
1	0	1
2	0	2
3	1	3
4	0	1
5	1	2

We start with a linear model as an approximation to the internal potential. Under this assumption, the internal potential can be expressed as follows (Brillinger (1988b))

$$U_t = \sum_{u=0}^{g_t - 1} a_u X_{t-u}, \tag{3.3.1}$$

where $g_t = \min\{\gamma_t, G\}$ and G is the index of the corresponding interval that contains the maximum time lag of interest. We usually set the maximum time lag of interest to be 1.0, since our main interest is in short time intervals, i.e. less than 1 s. Brillinger (1988a) uses γ_t in the upper limit to take into account the effect of the response process Y on itself. However, we truncate the sum to decrease the amount of computations, and because of our interest in short times intervals.

We arbitrarily set a response to occur at time 0, so that interval 0 contains a response, and $Y_0 = 1$, from which it follows that $\gamma_1 = 1$. That is, we assume that recording of the flashes and responses begins immediately after some initial response.

If we do not make such an assumption, then we have $\gamma_j = 0$ for all time interval indices j until a response occurs in some time interval. However, this is problematic because when $\gamma_j = 0$, it can be seen from (3.3.1) that $U_t = 0$, and there is no dependency on the stimuli until a response occurs.

A response occurs when

$$U_t + \varepsilon_t > \theta,$$

where θ is the threshold, and ε_t is internal noise, which we assume to have a symmetric distribution about 0. In our model, the threshold is fixed and internal noise is added to the internal potential U_t. This differs from Brillinger's model, where a random threshold is used; at time t the threshold is $\theta_t = \theta + \varepsilon_t$, and a neuron fires when $U_t > \theta + \varepsilon_t$. However, by restricting the distribution of the noise to be symmetric about zero, the form of the likelihood remains the same for both formulations, and only the interpretations differ. Both models have two sources of variation, one of them from whether or not a response occurs (neuron fires) given the threshold level. The other source of variation is derived from the random threshold in Brillinger's model, and from the internal noise in the total internal potential of our model.

We also look at an extension of the linear model, so that interaction terms are included. We take into account the fact that a response to a flash may be dependent on the occurrence of previous flashes that are close together in time with the present one. That is, we consider interactions of a flash at time t with the occurrence of any flashes that are up to d time units earlier, for some d. The parameter d is related to the d parameter in the parametric RT model of Braun et al. (2003). In that model, when two consecutive flashes occur within d time units apart, one of the two flashes is deleted (not responded to). For simplicity, in the threshold model, we set $d = rm$, for some integer r, since we are working with data that is discretized using lag m. Letting $Z_j = I\{\sum_{i=1}^{r} X_{j-i} > 0\}$, the internal potential can be expressed as

$$U_t = \sum_{u=0}^{g_t-1} a_u X_{t-u} + \sum_{u=0}^{z_t-1} b_u X_{t-u} Z_{t-u}, \tag{3.3.2}$$

where $z_t = \min\{\gamma_t, z\}$ is the maximum number of time intervals (prior to the current time interval) that are of interest for an interaction effect on the internal potential.

Thus, $z \leq G$.

For the threshold model of the simple RT experiment in which two flash types are presented, the notation is identical, but we must introduce additional notation to distinguish between the two flash types. We denote the discretized black and white flashes, by X^B and X^W, respectively. For our model without interactions the internal potential is now

$$U_t = \sum_{u=0}^{g_t-1} (a_u X^B_{t-u} + b_u X^W_{t-u}),$$

where, as before, $g_t = \min\{\gamma_t, G\}$ and G is the index of the corresponding interval that contains the maximum time of interest. In the simplest model $a_u = b_u$, which corresponds to the case in which black and white flashes lead to effects of the same magnitude on the internal potential; the effect of a black flash in increasing U_t is the same as the effect of a white flash in increasing U_t. Intuitively, this appears to be a reasonable assumption. Thus, we make the assumption of equal effects $a_u = b_u$, which allows us to ignore flash type so that the internal potential is identical to (3.3.1), and in the case of interactions the internal potential is (3.3.2).

3.3.1 Statistical Methods

We fit binomial models with probit (standard normal error distribution), and logit (logistic error distribution with location 0 and scale 1) link functions to simple RT data with one flash type by using the `glm` function of R (R Development Core Team (2006)). Our fitted model has the form

$$\hat{Y}_j = \begin{cases} 0, & \text{if } \hat{U}_j < \hat{\theta} \\ 1, & \text{if } \hat{U}_j > \hat{\theta} \end{cases},$$

where

$$\hat{U}_j = \sum_{u=0}^{g_j-1} \hat{a}_u X_{j-u},$$

in the case of no interactions, and

$$\hat{U}_j = \sum_{u=0}^{g_j-1} \hat{a}_u X_{j-u} + \sum_{u=0}^{z_j-1} \hat{b}_u X_{j-u} Z_{j-u},$$

in the case of inclusion of interaction terms.

In R we fit the model

$$\hat{Y}_j = \begin{cases} 0, & \text{if } \hat{Y}_j^* < 0 \\ 1, & \text{if } \hat{Y}_j^* > 0 \end{cases},$$

where in the case that interaction terms are included

$$\hat{Y}_j^* = \sum_{u=0}^{g_j-1} \hat{a}_u X_{j-u} + \sum_{u=0}^{z_j-1} \hat{b}_u X_{j-u} Z_{j-u} + \hat{\beta}$$

so that the estimate of the threshold $\hat{\theta}$ is the negative of the intercept estimate β.

Prior to fitting the model to real data, we fit the integrate-and-fire model using probit and logit links for data simulated from an integrate-and-fire model, and from a simple RT parametric model (Braun et al. (2003)). We make comparisons between the coefficient estimates a_t from the fitted threshold models, and the first-order kernel estimates $\hat{a}_1(t) = \frac{\hat{p}_{AB}(t)}{p_A} - \hat{p}_B = \hat{m}_{AB}(t) - \hat{p}_B$ discussed in Section 2.3.2; both $\hat{a}_t$ and $\hat{a}_1(t)$ give estimates of the average impulse response. When converting the simple RT data with one flash type to 0-1 time series we use a lag of $m = .07$.

The results obtained using the standard normal error distribution and using the logistic error distribution are virtually the same except for scaling. This gives us evidence that the fitting procedure is somewhat robust with respect to the assumed error distribution. A similar observation was made by Brillinger (1988a) for threshold models fit to neuron data. For this reason, we only consider the logit link in the analysis of simple RT data with two flash types, and in the remaining chapters.

For the simple RT data with two flash types we consider logit link functions and fit binomial models, as in the case of one flash type. We use a lag of $m = 0.05$, and fit models to nine real data sets, each having a different stimulus rate; the rates of flashes per second are 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 2.0, 4.0, and 8.0.

3.3.2 Diagnostics

As in Brillinger (1988b), we can assess goodness of fit of either model by examining plots of the empirical probability of responding against the corresponding predictor $\hat{U}_t$. The probability of a response will increase with U_t since a response occurs when

the predictor, with some random noise added to it, crosses the threshold θ; the larger the value of U_t, the higher the probability of a response. Using a grid of u values that span the range of the $\hat{U}_t$ the empirical probability of a response is the proportion of responses that occur when $\hat{U}_t$ is in a small interval near each u. For small h, an estimated probability of a response is given by

$$\frac{\sharp\{Y_t = 1 \text{ with } u - h < \hat{U}_t < u + h\}}{\sharp\{t \text{ with } u - h < \hat{U}_t < u + h\}}.$$

Under the assumption that the model is correct, the probability of a response is

$$P(U_t + \varepsilon_t > \theta) = P(\varepsilon_t > \theta - \hat{U}_t).$$

For a model that fits the data well, the empirical probabilities should follow the curve of the fitted probabilities.

3.3.3 Application to Integrate-and-fire Simulations

As a means of studying the goodness of fit and robustness of the integrate-and-fire (i-f) model we apply it to data simulated from such a model. We consider standard normal and standard logistic error distributions, corresponding to probit and logit links, respectively. To study the goodness of fit, we fit a model in which the link function matches the error distribution of the simulation. We study the robustness by fitting a model with a link function different from the error distribution of the simulation.

As a simple case, we first study an i-f model in which only the first two coefficients, a_0 and a_1, are non-zero. When the probit link i-f model is fit to i-f simulated data with a standard normal noise distribution, the fit is very good. In the diagnostic plot (not shown) the empirical probability points follow the theoretical curve with the exception of one outlier. The first two coefficients are very close to the true values used for the simulation and are clearly non-zero. Although $a_2 = 0$ for the simulation, in this fitted model, it is significantly different from 0 with an estimate of 0.41 and a standard error of .15.

When we fit an i-f model with a logit link to the i-f simulated data with normal noise, we find that the first three coefficients are significantly different from 0, as in the fit using the probit link. However, the diagnostic indicates a poor fit, and the estimates are far from the true values.

A plot comparing the two coefficient estimates obtained from fits with the two links, and with $a_1(u)$ is given in Figure 3.2. The graph corresponding to the model fit with the logit link appears to have the same features as the one for the probit link, but on a different scale. It can also be seen that where the coefficients are actually zero all three estimation curves fluctuate around zero, with the logit link model estimates having the largest variability.

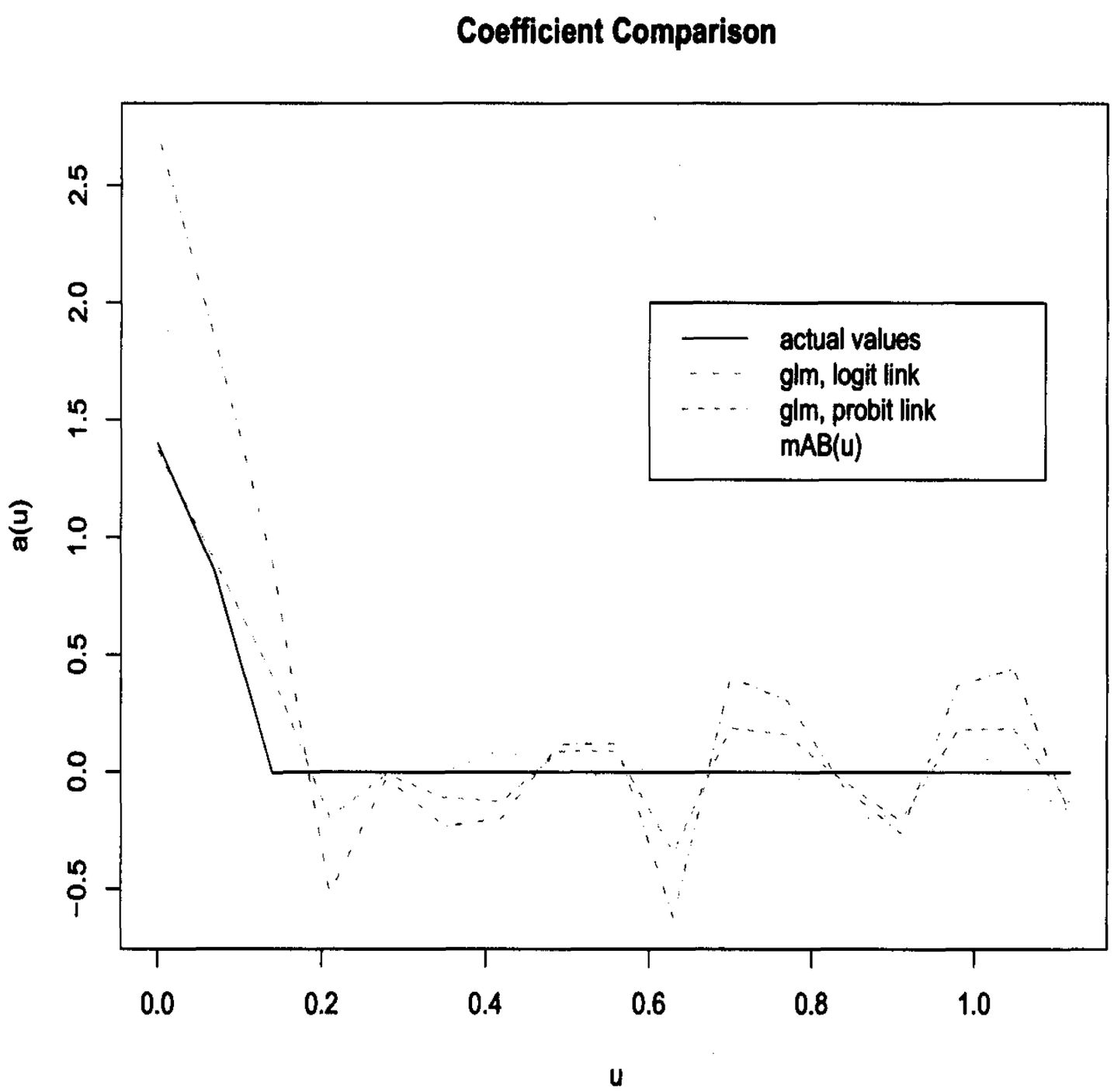


Figure 3.2: Comparison plots for linear integrate-and-fire models fit to a simulated data set from an i-f model with normal error distribution and having only the first two coefficients nonzero.

Similarly, we simulate data from an i-f model with logistic noise, and fit i-f models with probit and logit link functions to the simulated data. The diagnostic plot (not shown) for the logit link fit indicates that a model with a logit link is better than one with a probit link. As expected, the estimates based on the logit link are very close to the actual values.

A similar comparison was done for simulated data from an i-f model in which $a_0, \ldots, a_9$ are nonzero. As in the previous simulations, for the normal noise simulation the probit link model yields estimates with a smaller bias than the logit link model, and in the same sense the logit link model is better than the probit link model for a logistic noise simulation. However, unlike the case with only two non-zero coefficients, the diagnostic plots are very similar for the two link functions. That is, as the number of non-zero coefficients increase, there are fewer differences in the diagnostics for models with different link functions. Furthermore, the fits for the data with more non-zero coefficients (e.g. nine) are better than those fit to data with few nonzero coefficients (e.g. two). Diagnostic plots for the two models fit to the simulation with a normal error distribution and nine non-zero coefficients are given in Figure 3.3.

When an i-f model with interaction terms is fit to the i-f simulated data considered above, as expected, the interaction terms are not discernible from zero.

In summary, we find that when the link function of the fitted model is the same as the error distribution of the simulation, the diagnostics indicate a very good fit, and the coefficient estimates are near the true values. When we fit a model with a link function different from the error distribution of the simulation, we find that the robustness of the model rises as the number of nonzero coefficients in the simulation increases. That is, when the number of coefficients increases, the diagnostics indicate a good fit regardless of the link function employed. A different scaling in the graphs of the coefficients is the only observable difference in fits of models with different link functions.

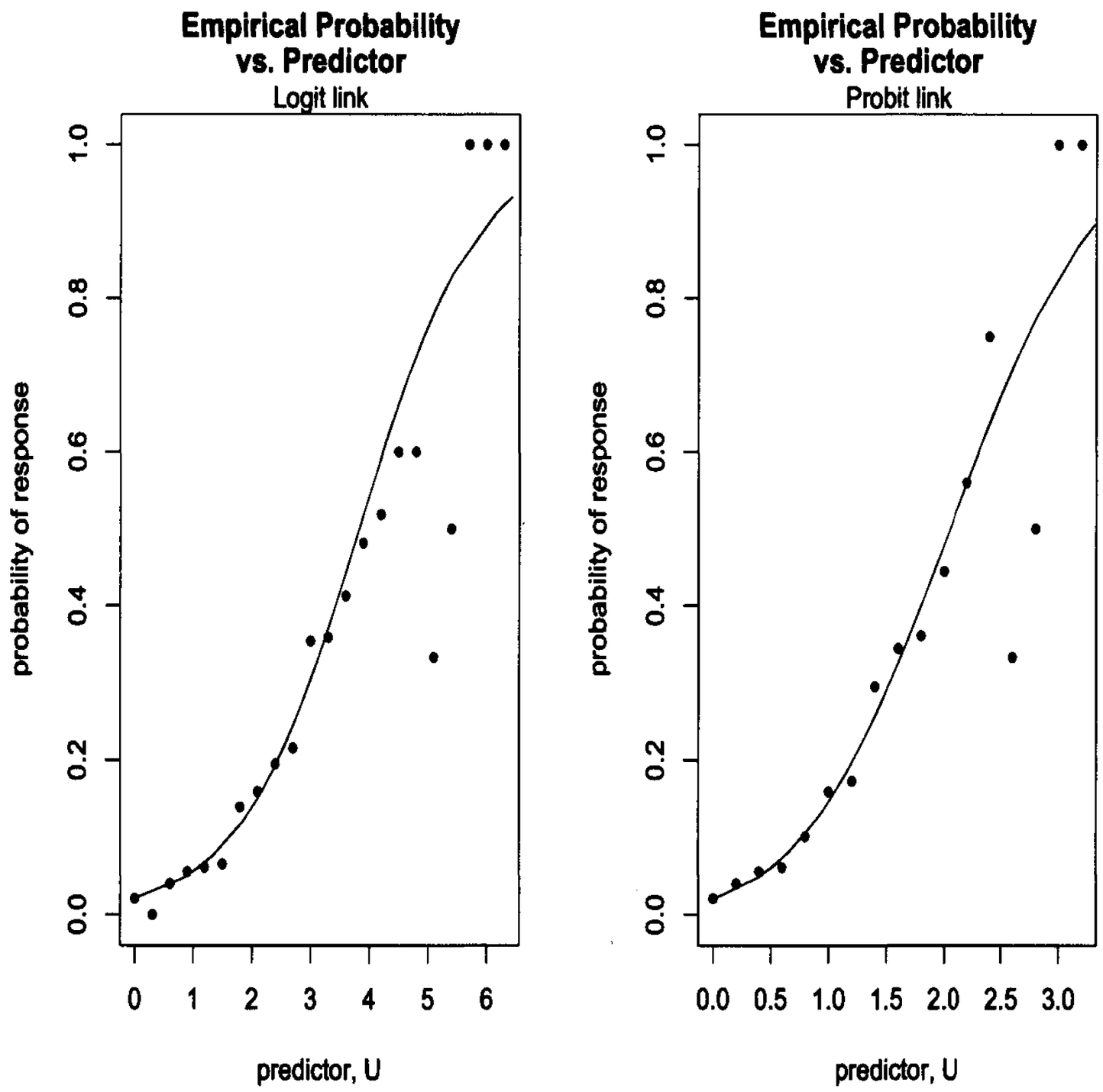


Figure 3.3: Diagnostic plots for simulated data from an i-f model with a normal error distribution and having the first nine coefficients non-zero. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

3.3.4 Application to RT thinning model simulations

We fit the integrate-and-fire model with probit and logit links to simulations that we generated from the parametric simple RT model of Braun et al. (2003) with various parameter settings.

For data that we simulated from the parametric RT model with $d = 0$, the diagnostic plots (for logit and probit) indicate that the i-f fit without interactions is excellent for $U_t < \hat{\theta}$. For larger U_t, most of the empirical probabilities do not follow the curve as well, and many fall below the fitted probabilities.

As d increases, the fit of the i-f model without interactions worsens, according

to the diagnostic plots. For all d, when $U_t \leq 0$ the fit is excellent. However, the corresponding probabilities are near 0 and there are very few responses so such a good fit is expected; under the i-f model there are no responses when $U_t < \theta$ so the probability of a response is near 0. For $d > 0$, when $0 < U_t < \hat{\theta}$ the i-f model fits well, but tends to underestimate the probability of a response, and it overestimates the probability of a response when $U_t > \hat{\theta}$. We expect such results since the i-f model will predict a low probability of a response when $U_t < \theta$, and as U_t increases above θ the response probability increases. There is likely more dependence among the responses and their history of flashes than the form of the i-f model considered here.

It is difficult to see any effect that σ and p have on the fit. There is a tendency for the fit to be slightly better when $\sigma = .08$ than when $\sigma = .12$, but the model still does not fit well. There does not appear to be a difference in the fit when we simulate the data under either of the two thinning assumptions, T.1 and T.2. The estimates of the threshold θ tend to be around 2.0 for the probit link model, and around 3.6 for the logit link model.

In the i-f model without interactions a response occurs when $U_t+\varepsilon_t > \theta$, where U_t is a function of the indicators of flash occurrences from the time after the last response up to the current time. Since the coefficients of the indicators tend to be positive, when there are no responses to many flashes the value of U_t (plus random noise) increases and is eventually greater than the threshold θ, resulting in a response. The inclusion of interaction terms yields a better fit, with significant coefficients of interaction terms being negative; interaction terms decrease U_t resulting in a lower probability of U_t crossing the threshold, and a response occurring. Negative interaction coefficients are indicative of inhibition among the flashes, which we know is present when $d > 0$ in the parametric model.

When the i-f model with interactions is fit to RT data simulated from a model with $d = 0$, the interaction terms are not usually needed, and when some are needed the diagnostic indicates an improvement over i-f models without interaction terms. In the i-f model, $d = .21$ is usually used since there appears to be no improvement in using a smaller d such as $d = .14$ or $d = .07$. Similarities in the fits of the models

with and without interactions lessen as σ and p increase. The models with interaction terms fit quite well, except for an occasional outlier for large U.

When $d > 0$, fits involving the i-f models with interactions are an improvement over those without interaction terms, sometimes giving an excellent fit. The fit tends to worsen as σ increases from .08 to .12, and there does not appear to be a difference in the fits for data generated under thinning assumptions T.1 and T.2. The fits under the probit model and the logit model are very similar, as illustrated in Figure 3.4. In that figure logit and probit models with $d = .21$ are fit to simulated RT data under thinning assumption T.2 with parameters $d = .2, \sigma = .08, p = .15$, and $\mu = .4$.

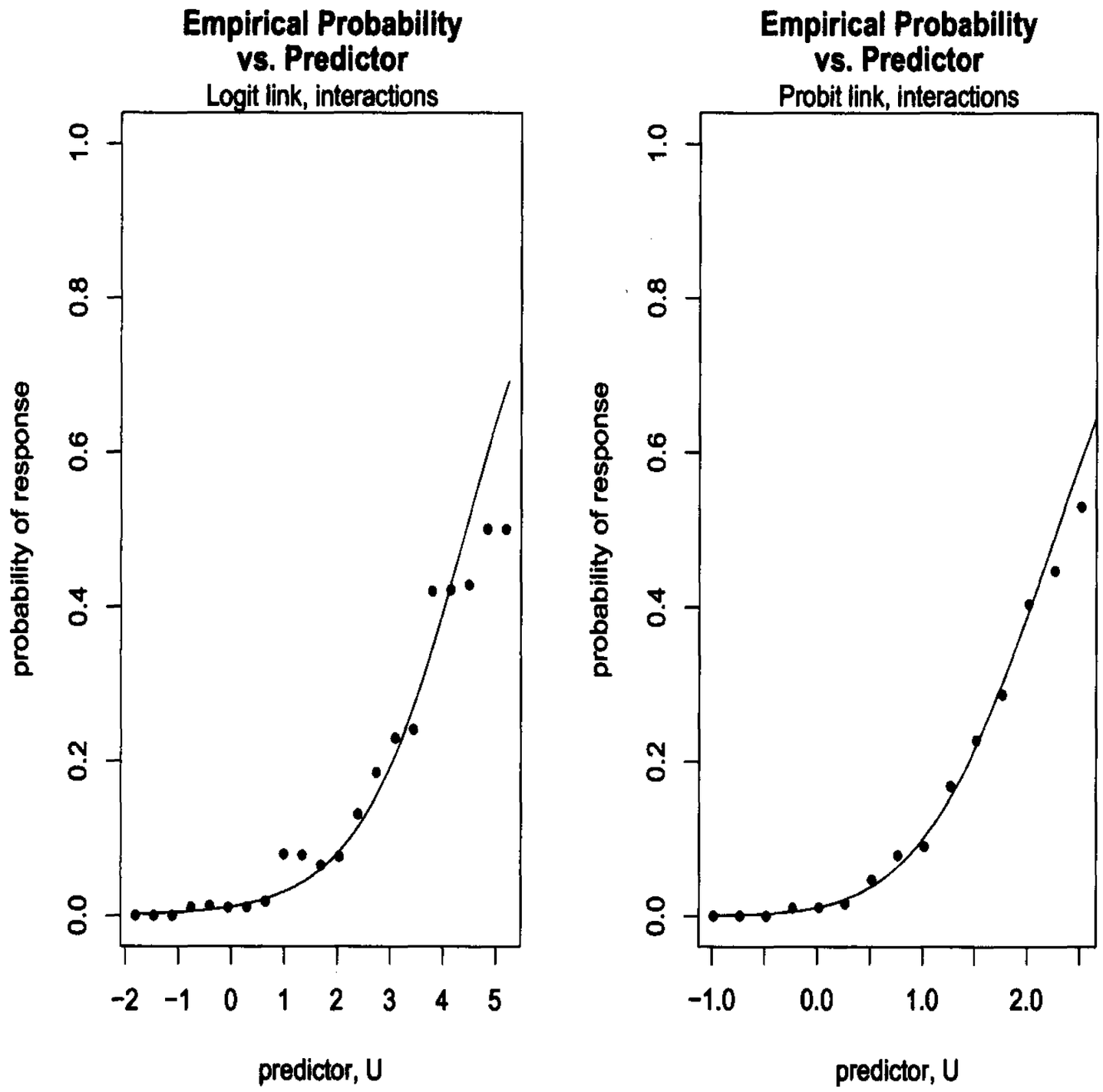


Figure 3.4: Diagnostic plots for integrate-and-fire models with interactions (d=.21) fit to simulated RT data under thinning assumption T.2 with parameters $d = .2, \sigma = .08, p = .15, \mu = .4$.The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

3.3.5 Application to Simple RT Data - One Flash Type

The fits for the 12 RT data sets with flash rate 1.0 s are similar, and the θ estimates are provided in Table 3.4. Under the logit i-f linear model the estimate of θ across the 12 data sets has sample mean 4.09, standard error 0.177, and under the logit model with interactions, the mean and standard error are 4.37 and 0.222, respectively. Under the probit model without interactions the estimates have mean 2.18 and standard error .077, while for the model with interactions the mean and standard error are 2.26 and .088, respectively. These results give an indication that the value of θ is very similar among the data sets.

Table 3.4: Threshold estimates $\hat{\theta}$ for RT data with flash rate $p_A = 1.0$.

	without interactions		with interactions	
data set	$\hat{\theta}$, logit	$\hat{\theta}$, probit	$\hat{\theta}$, logit	$\hat{\theta}$, probit
1	4.36	2.32	4.65	2.40
2	4.35	2.29	4.53	2.33
3	4.07	2.17	4.18	2.18
4	3.96	2.13	4.33	2.23
5	3.91	2.11	4.20	2.19
6	4.28	2.24	4.49	2.30
7	3.97	2.14	4.19	2.20
8	4.20	2.24	4.73	2.37
9	4.13	2.21	4.58	2.34
10	4.00	2.16	4.34	2.26
11	3.80	2.05	3.98	2.10
12	4.05	2.19	4.32	2.26

As a comparison, for the first RT data set, the coefficient estimates for the model without interactions are plotted along with the first order kernel estimates $a_1(u) = m_{AB}(u) - p_B$, in Figure 3.5. All three estimates peak in the same location, but are of different magnitudes, and as observed in other examples, the estimation curves

corresponding to the probit and logit link models are identical, except for a a different scale. For each of the RT data sets the diagnostic plots indicate a poor fit for $U_t > \hat{\theta}$.

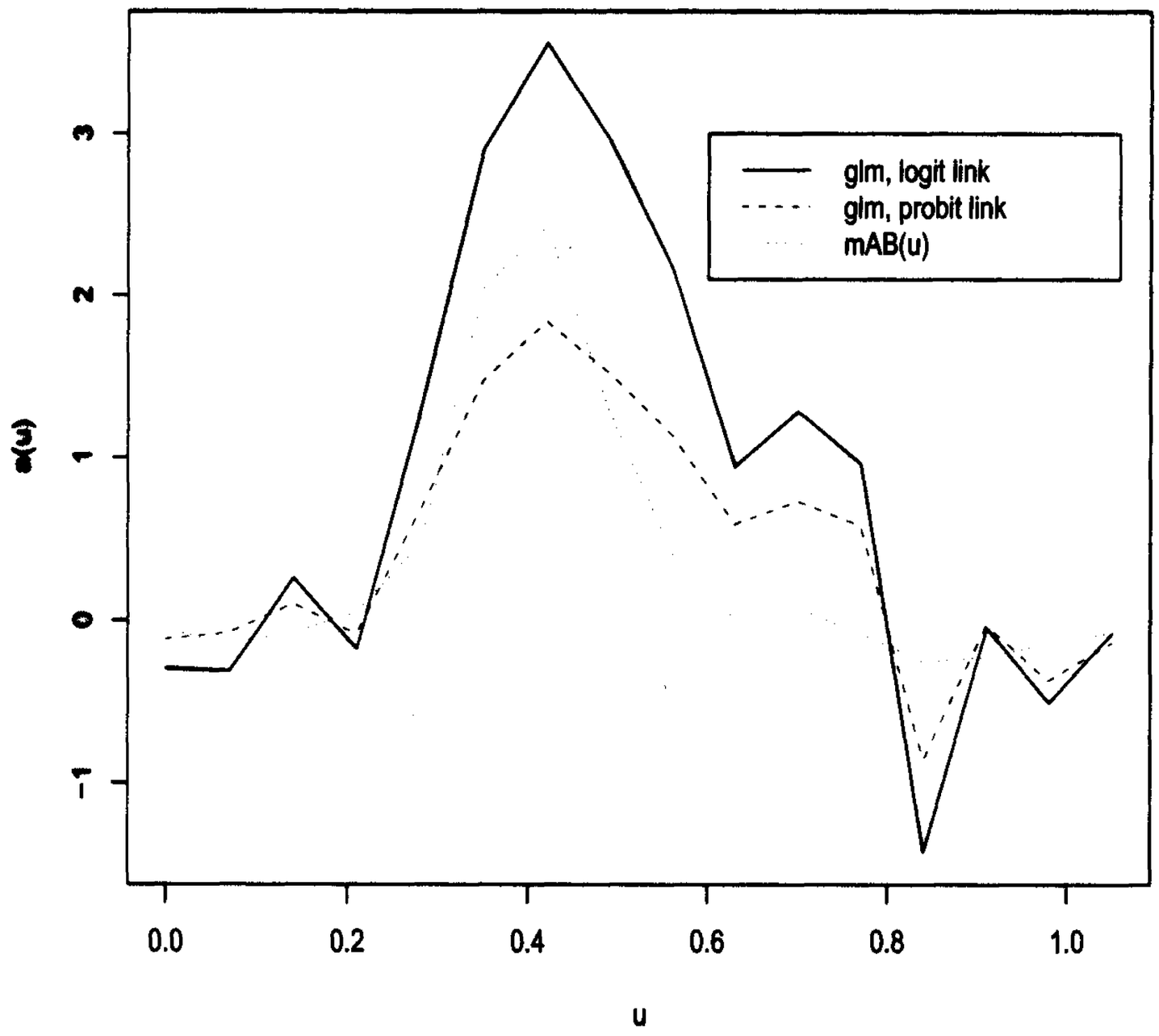


Figure 3.5: Comparison plots for linear integrate-and-fire models fit to the first RT data set with flash rate $p_A = 1.0$.

Diagnostic plots of the i-f model with interactions fit to the RT data reveal an improvement over the models without interaction terms. In some cases, such as the second data set, the model fits quite well. Diagnostic plots for the second data set are provided in Figure 3.6; the logit model with interactions is an excellent fit with the exception of one outlier.

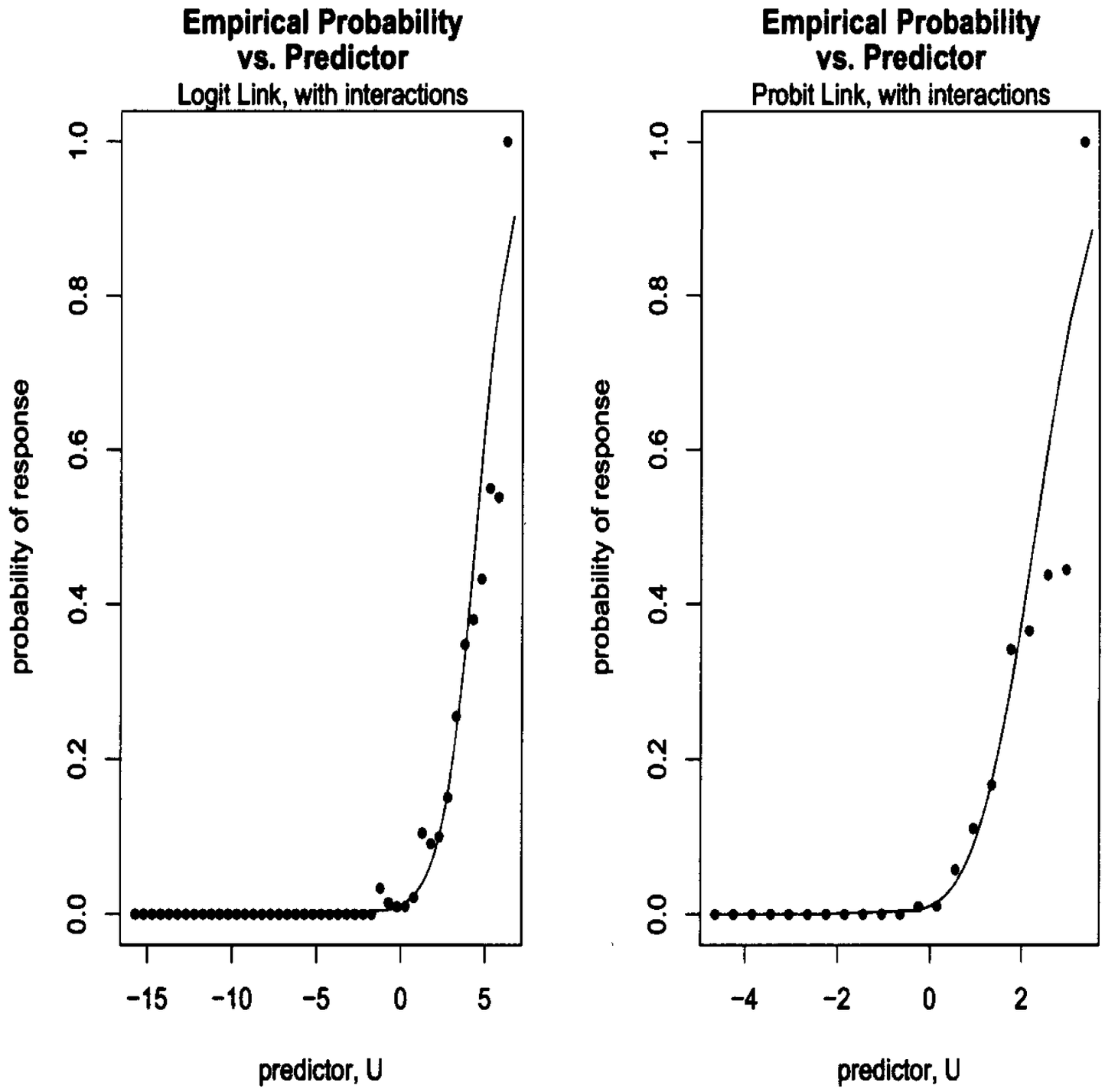


Figure 3.6: Diagnostic plots for linear integrate-and-fire models with interactions fit to the second RT data set with flash rate $p_A = 1.0$. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

3.3.6 Application to Pooled Simple RT Data - One Flash Type

Since the 12 RT data sets are based on the same person, we assume that they should have the same parameters for the i-f model. In the previous subsection, where each data set is fit individually, it appears that the threshold θ may be the same across the data sets. In order to get one set of estimates for the 12 data sets we pool them into one large data set. The RT data is pooled in a simple manner by concatenating the 12 data sets. Let M_i and N_i denote the number of flashes and the number of

responses for data set i, respectively. Then the pooled RT data is

$$A_{11}, A_{12}, \ldots, A_{1M_1}, A_{21} + A_{1M_1}, \ldots, A_{2M_2} + A_{1M_1}, A_{31} + A_{2M_2} + A_{1M_1}, \ldots, \sum_{i=1}^{12} A_{iM_i}$$

and

$$B_{11}, B_{12}, \ldots, B_{1N_1}, B_{21}+A_{1M_1}, \ldots, B_{2N_2}+A_{1M_1}, B_{31}+A_{2M_2}+A_{1M_1}, \ldots, B_{12M_{12}}+\sum_{i=1}^{12} A_{iM_i}.$$

When converting the data to 0-1 time series a lag of $m = .07$ is used and only coefficients corresponding to lags less than $1.0s$ are included in the model.

For both of the i-f models, all of the coefficients corresponding to intervals in $[0.21, 0.77]$ are discernible from 0. This is an indication that button presses (responses) tend to occur between 0.21 s and 0.77 s after a flash. For the model with interactions the interaction coefficients that are discernible from zero correspond to intervals in $[0.21, 0.63]$, indicating that flashes occurring 0.21 to 0.63 s before the current flash influence the occurrence of a response to the current flash. The coefficients a_u for the logit and probit link models with interactions are displayed in Figure 3.7.

The threshold estimates $\hat{\theta}$ are given in Table 3.5. Recall that when the internal potential rises above the threshold, a response occurs. The diagnostic plots indicate

Table 3.5: Threshold estimates $\hat{\theta}$ for pooled RT data with flash rate $p_A = 1.0$

	logit $\hat{\theta}$	probit $\hat{\theta}$
without interactions	3.98	2.15
with interactions	4.29	2.23

that the linear i-f models are a poor fit to the pooled RT data, and are provided in Figure 3.8. However, inclusion of interactions yield a better fitting model, as can be seen in Figure 3.9.

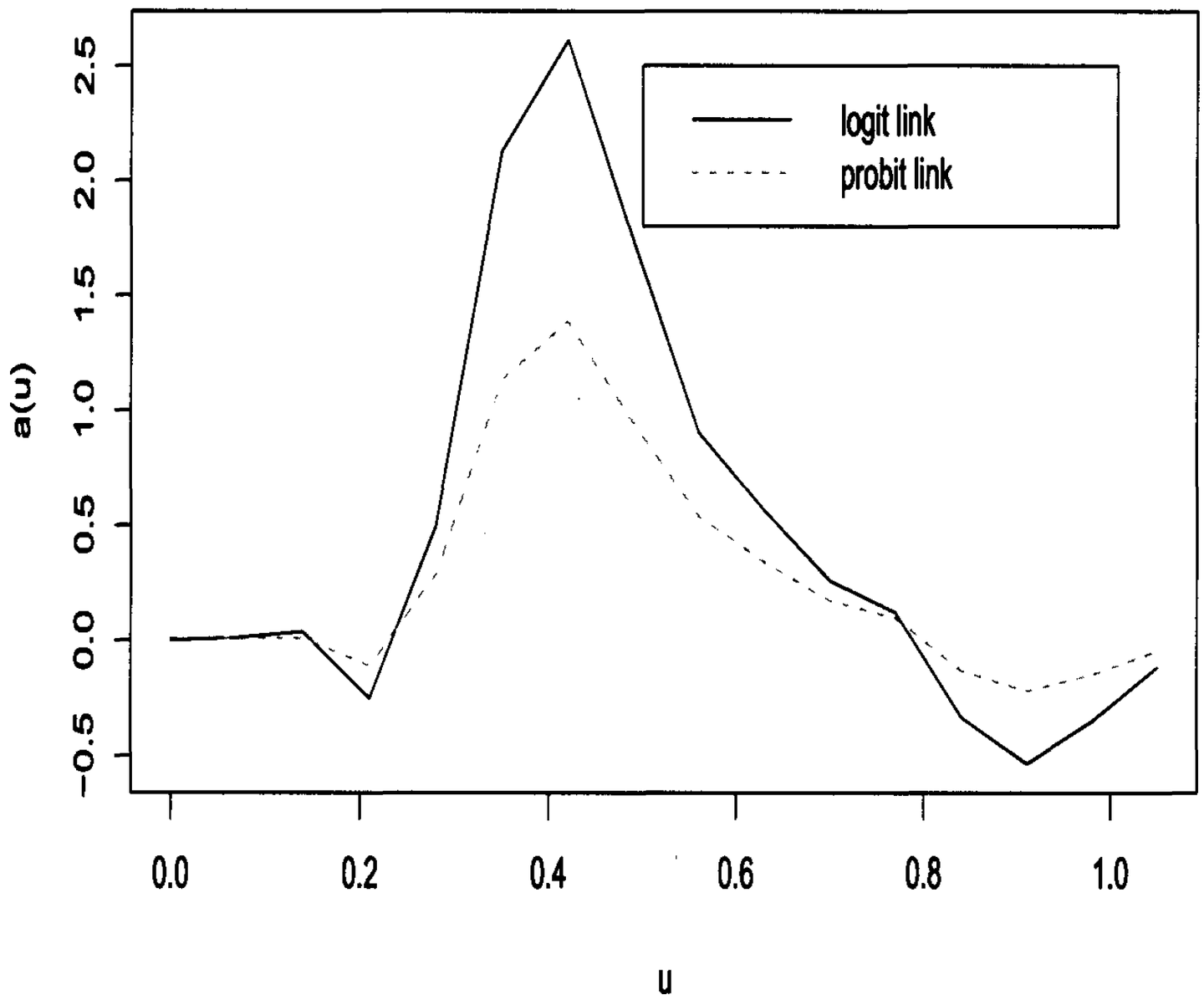
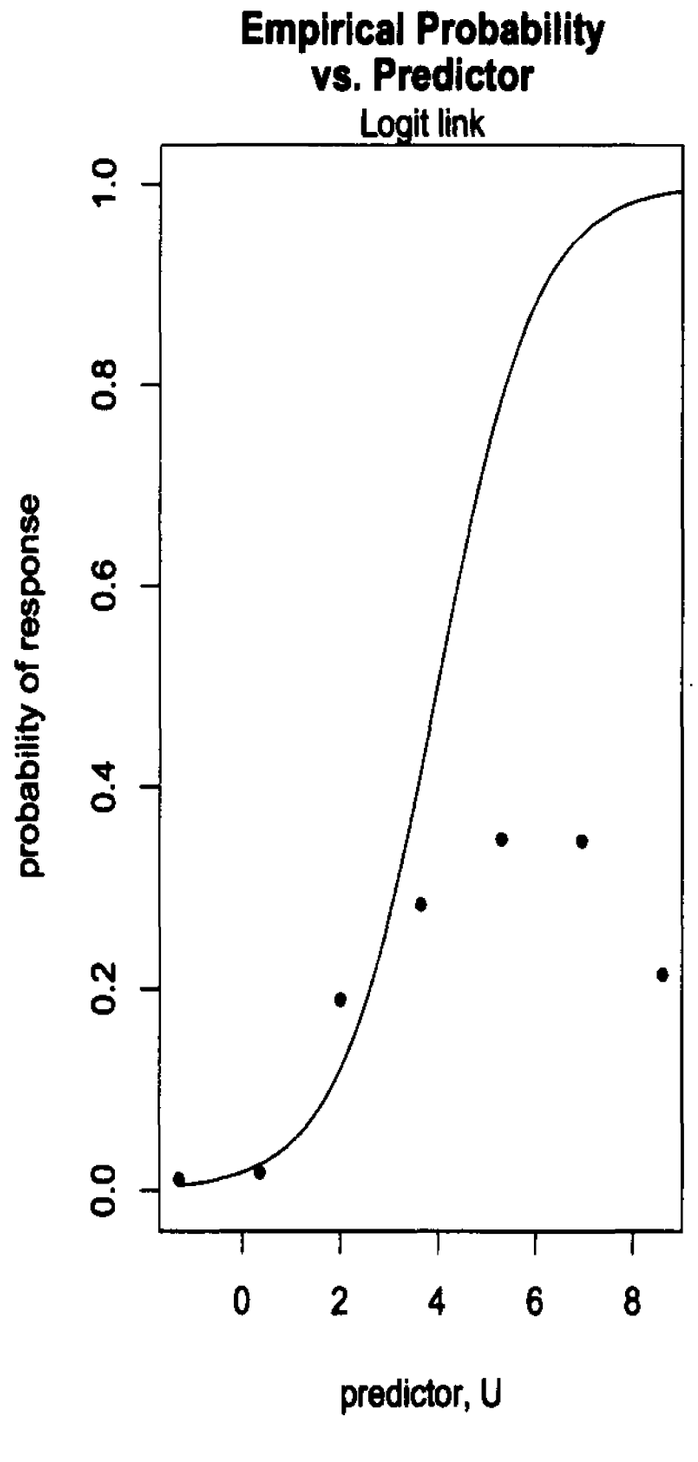


Figure 3.7: Impulse responses for threshold models with interactions and logit and probit links, respectively, fit to pooled RT data with $p_A = 1.0$.

3.3.7 Application to Simple RT (two flash types)

Following the belief that data resulting from different runs of an experiment on the same individual can be fit by identical models, with differences due only to randomness between trials, our analysis is done on pooled data. Models with interactions are fit to nine pooled simple RT data sets, each having a different stimulus rate, and within each data set black and white flashes are presented at equal rates. Each pooled data set is based on eight experimental runs consisting of a total of 100 flashes each. As discussed in Section 3.3.1, we only fit models using the logit link, due to a difference in graph scaling being the only difference in the fits from different link functions. In the discretization of the data we use a lag of $m = 0.05$.



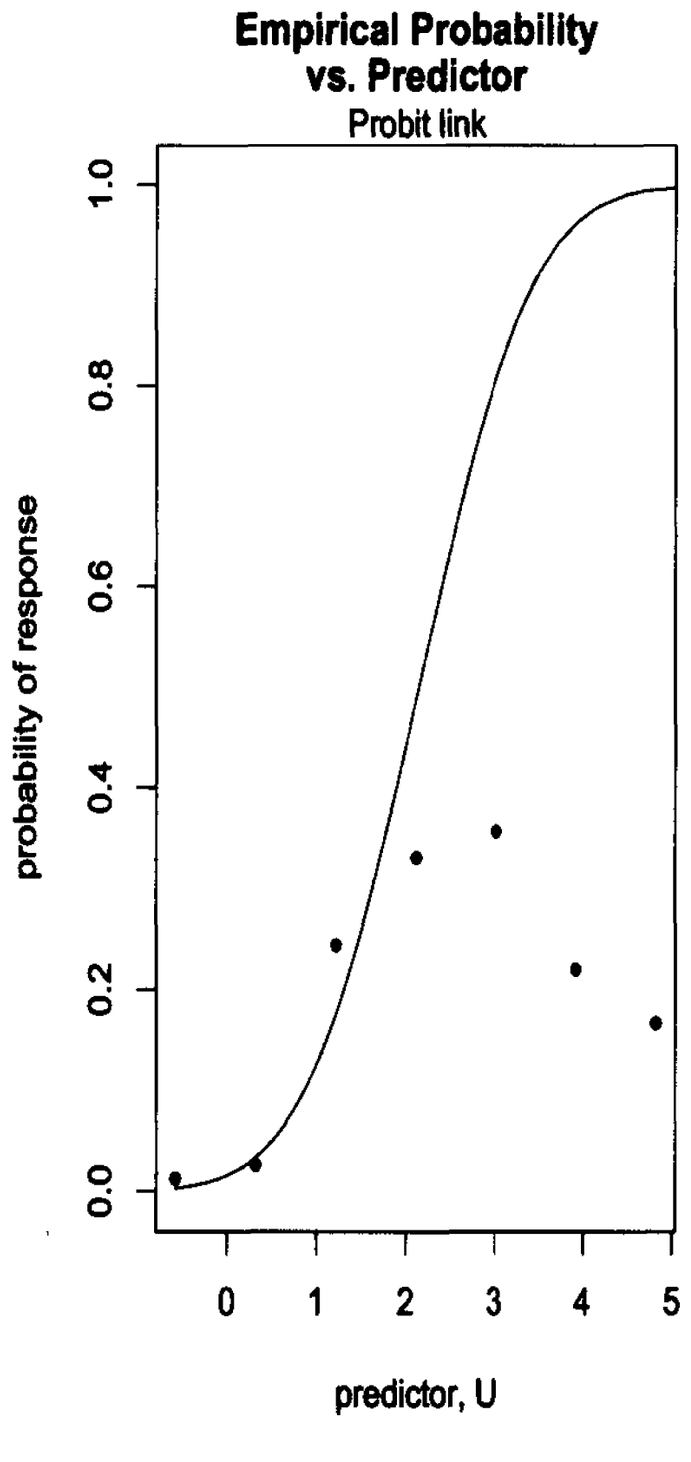


Figure 3.8: Diagnostic plots for linear integrate-and-fire models without interaction terms, and with logit and probit links, respectively, fit to pooled RT data with $p_A = 1.0$. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

Table 3.6 displays the threshold estimates, $\hat{\theta}$ for a model that fits the data reasonably well, based on the diagnostic plots. The "best-fitting" model varies for each of the data sets, with different d parameters, and different numbers of interaction terms. The thresholds tend to decrease with the stimulus rate.

Plots of the impulse responses, as well as the impulse response peaks as a function of stimulus rate, are provided in Figures 3.10 and 3.11. The peaks of the response impulses decrease with the stimulus rate. This is consistent with intuition. With a slower stimulus rate one is less likely to miss a stimulus. A higher peak is an indication of a larger frequency of responses at the lag for which the peak occurs. Examples

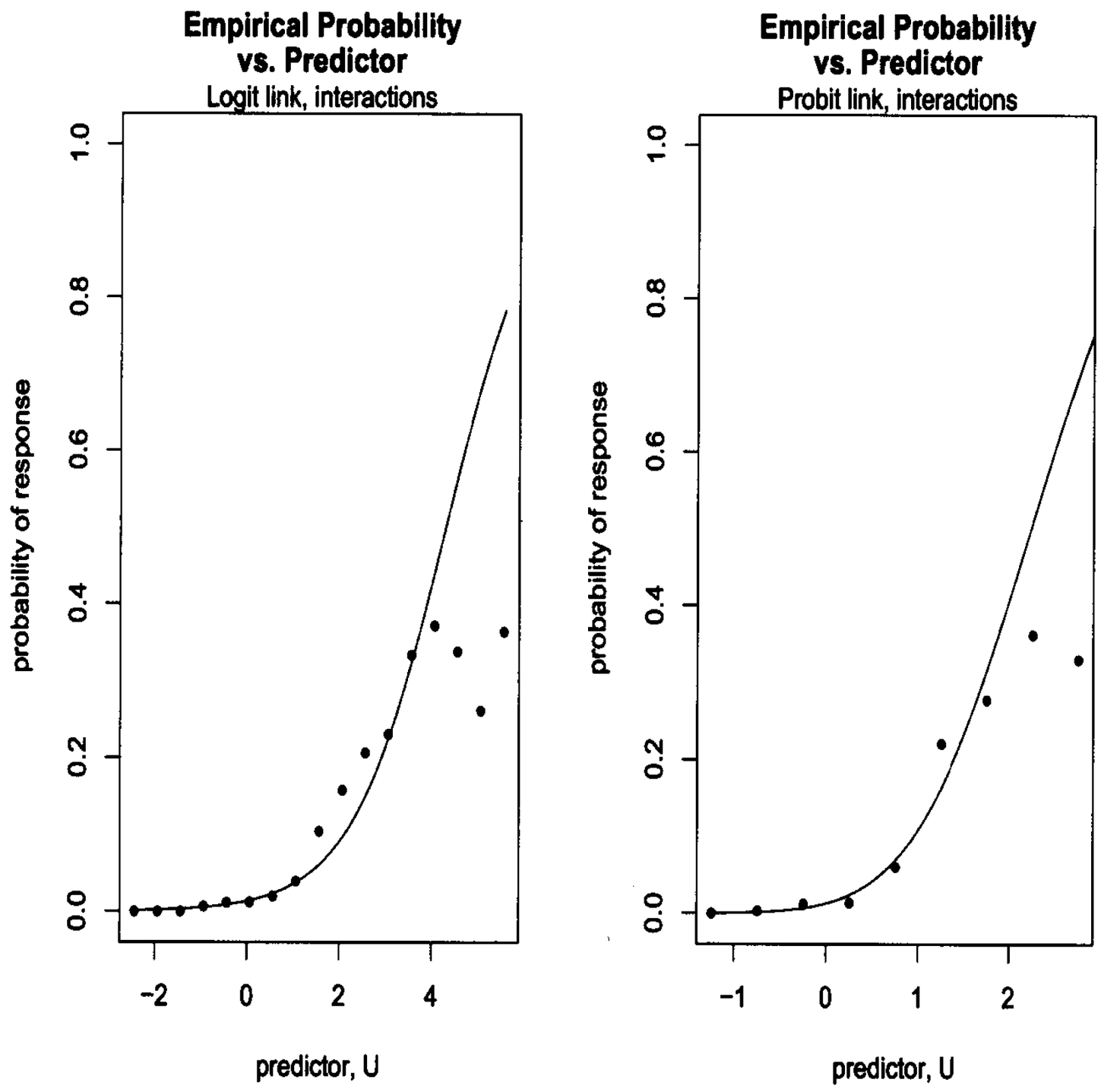


Figure 3.9: Diagnostic plots for linear integrate-and-fire models with interactions, and logit and probit links, respectively, fit to pooled RT data with $p_A = 1.0$. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

of diagnostic plots for four of the pooled data sets are provided in Figure 3.12, from which it can be seen that the model fits the data quite well.

3.4 Random Effects Threshold Model

We fit a random effects threshold model to the 12 sets of simple RT data with rate 1.0. As an alternative to pooling simple RT data sets with the same rate, we treat each data set as a realization of observations that could result from a single i-f model and consider any differences among the data to be due to randomness. Doing so

Table 3.6: Threshold estimates $\hat{\theta}$ for pooled simple RT data from experiments run at various flash rates. A logit link is used.

rate	$\hat{\theta}$
0.4	5.86
0.6	5.15
0.8	5.80
1.0	4.62
1.2	4.14
1.4	4.00
2.0	3.30
4.0	2.56
8.0	2.28

maintains the idea that the i-f model parameters should be the same for the twelve data sets. We fit a mixed effects model with the threshold and coefficients as fixed effects, and a random intercept with data set number as a grouping variable. Our model with interactions has the form:

$$Y_t = \begin{cases} 0, & \text{if } U_t < \theta \\ 1, & \text{if } U_t > \theta \end{cases},$$

where

$$U_t = \sum_{i=1}^{n} \sum_{u=0}^{g_t-1} a_u X_{i,t-u} + \sum_{i=1}^{n} \sum_{u=0}^{z_t-1} b_u X_{i,t-u} Z_{i,t-u} + \tau + \varepsilon_t,$$

where τ and ε_t, $t = 1, \ldots$ are independently distributed Normal random variables with mean zero and respective variances σ_τ^2 and σ^2, n is the number of data sets included in the model, and $X_{i,t}$ and $Z_{i,t}$ correspond to X_t and Z_t for the ith RT data set.

The random threshold model is fit in R using the function `glmmPQL` from the `MASS` library (Ripley (2006)). This function fits a generalized linear mixed model with multivariate normal random effects, using Penalized Quasi-Likelihood. We fit

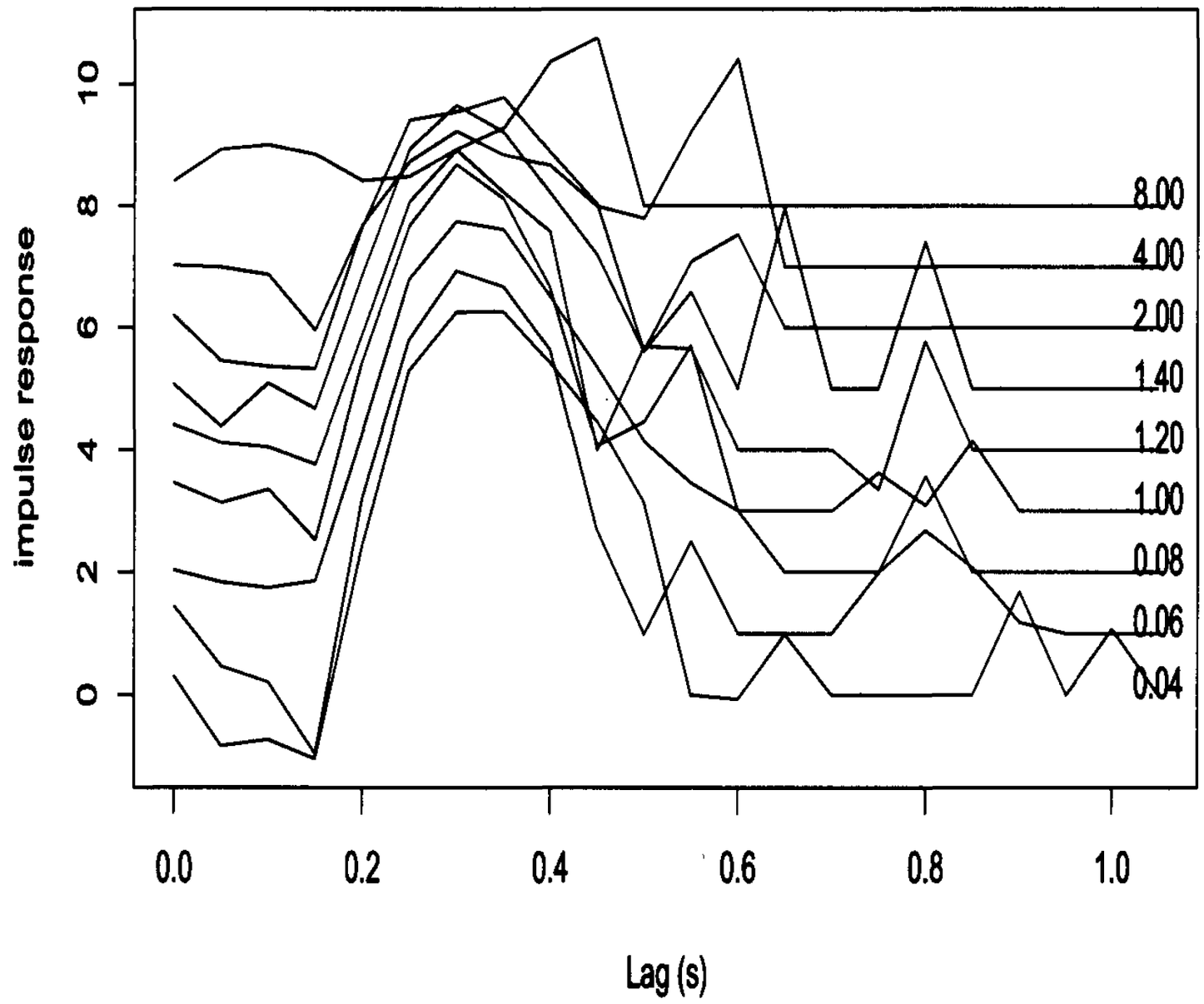


Figure 3.10: Plots of the stimulus linear filters for each of the 9 data sets. Curves are labeled by the stimulus rate (s). For ease of illustration, each curve is shifted upwards by one unit.

a model with logit link to the 12 data sets, and the impulse response is displayed in Figure 3.13, while the diagnostic plot is provided in Figure 3.14. In comparison with Figure 3.7, the impulse response from the model for pooled data, the same characteristics of a peak near 0.4, and a dip to the left of the peak appear in both graphs. Overall the results are the same, with the main difference between the two impulses being that the one resulting from the random effects model is slightly wider with a higher peak. The estimate of the threshold θ is 4.66, which is slightly higher than the estimate based on pooled data, $\hat{\theta} = 4.29$ (see Table 3.5).

The standard error of the random effects coefficient for data set number is 0.1201, indicating that there is not much variability among the data sets. In addition to the

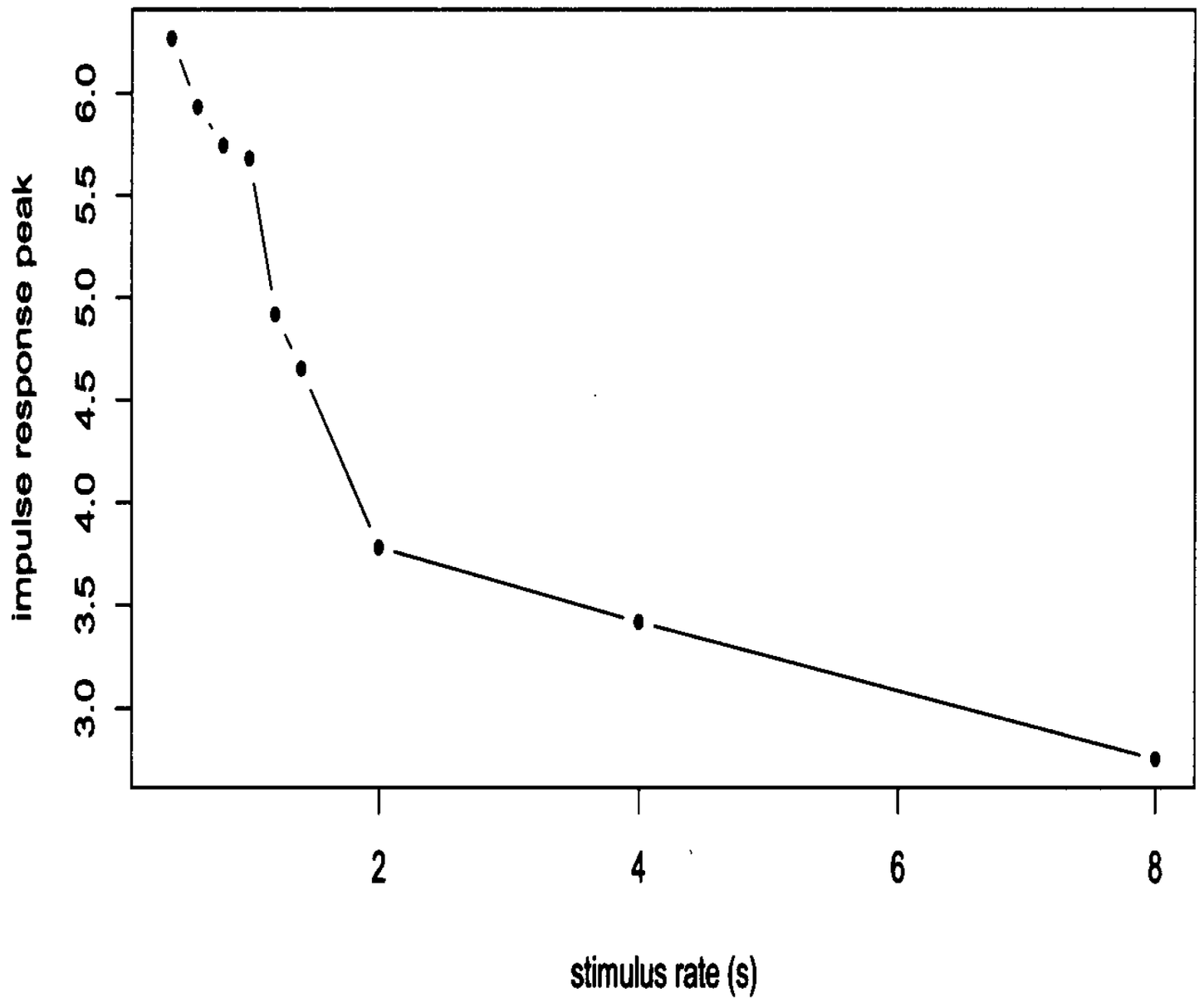


Figure 3.11: Plot of the impulse response peak height as a function of stimulus rate.

similarities in the impulse and threshold estimates, this gives us evidence that the data sets ares similar enough for the analysis to be done using simple pooling as in Section 3.3.6.

3.5 Discussion

In this chapter we continued to explore the parametric model of Braun et al. (2003), and introduced a threshold model for simple RT. Regarding the parametric model, details of the covariance approximation required for the computation of a test statistic for hypotheses involving the thinning mechanism are provided, as well as an alternative parameter estimation method involving nonlinear regression. The alternative

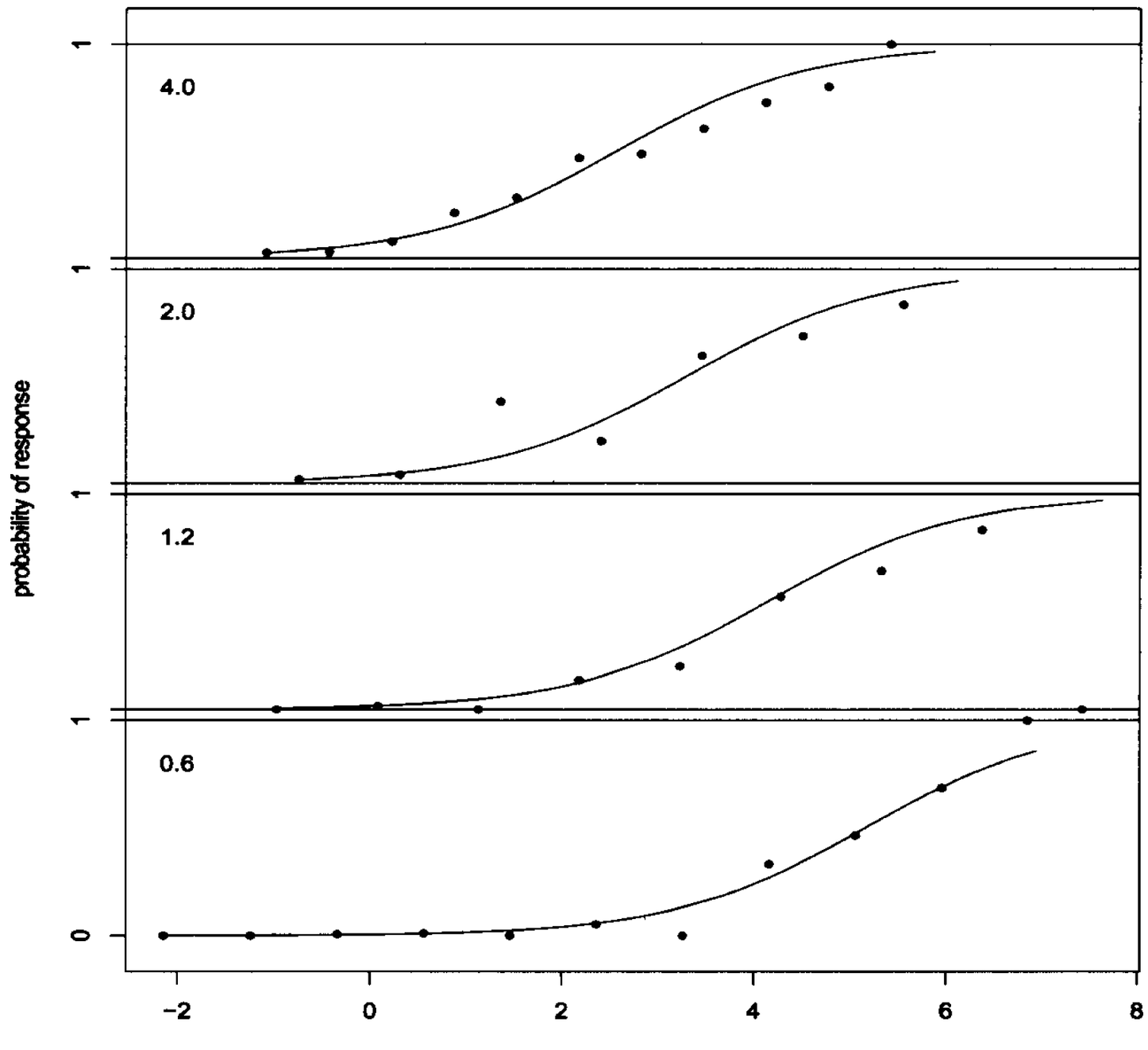


Figure 3.12: Diagnostic plots for the data sets with stimulus rates 0.6,1.2,2.0 and 4.0, as indicated in the top left corner of each plot. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

estimation method can be implemented when there are at least four simple RT data sets, each with a different flash rate, and it is assumed that the parameters are the same across the data sets. We found, based on a simulation study, that in terms of bias, variance, and MSE, nonlinear regression and the original parameter estimation method tend to have negligible differences in performance. Therefore, due to the advantages in using nonlinear regression we prefer this alternative method when it is reasonable to assume that the data sets come from the same population.

A threshold model for simple RT was introduced and applied to both simulated and real data. Fitting such models allowed us to obtain estimates of the threshold for each response type, as well as estimate certain intensities relating the stimuli to

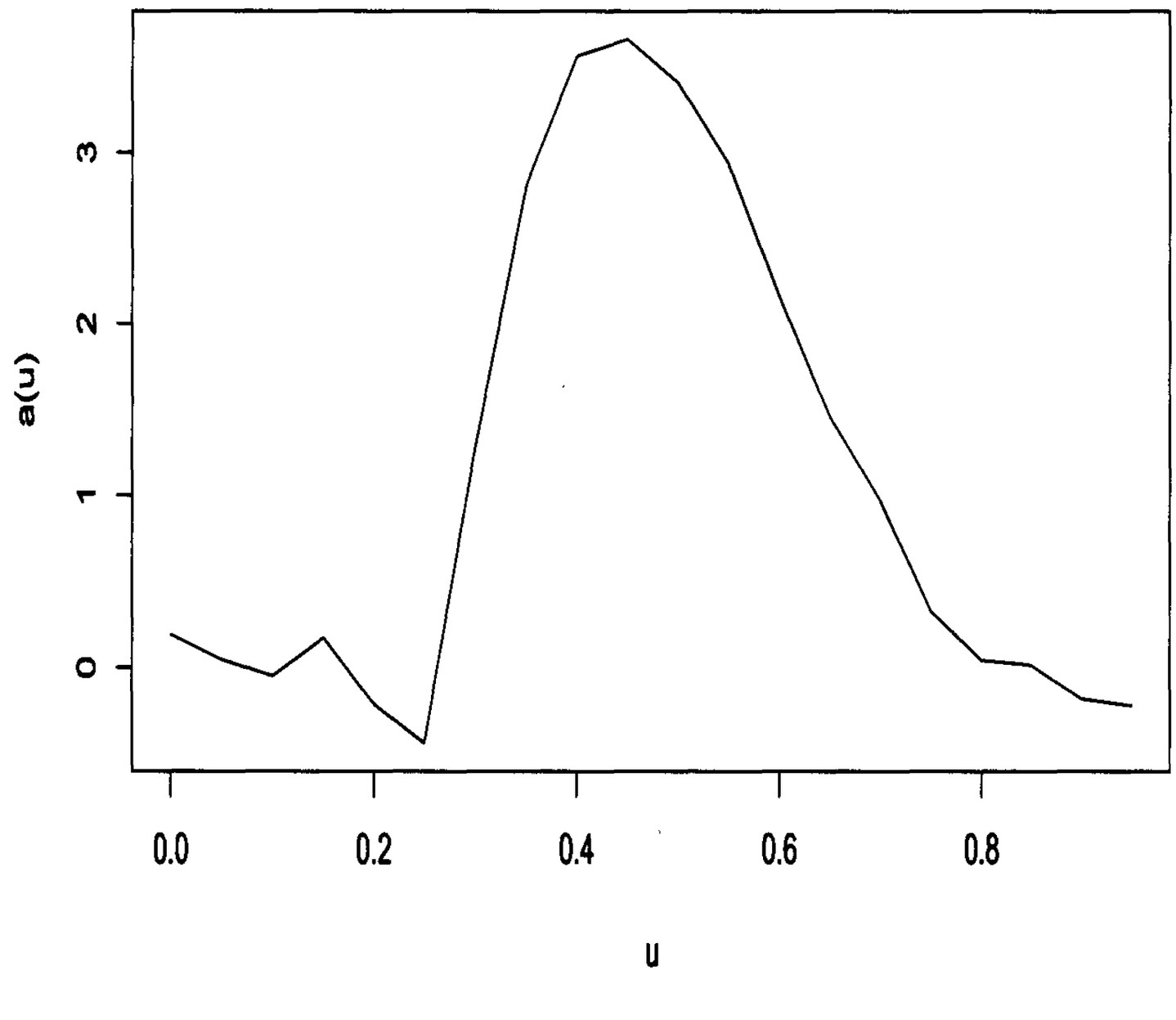


Figure 3.13: Impulse response for a random effects threshold model with logit link fit to 12 runs of simple RT data with flash rate $p_A = 1.0$.

the responses. Threshold estimates for the real data sets with different flash rates suggest that as the flash rate increases, a lower rod current is required for a response to occur. We fit models to data resulting from individual experimental runs, and to pooled data. Data with the same flash rates were pooled based on our assumption that since the data are the result of experiments run on the same person, the threshold and coefficients should be similar.

As an alternative to pooling the data to get a single set of parameter estimates, we proposed a random threshold model. Such a model incorporates variation due to different experimental runs, while maintaining the assumption that parameters are the same across runs of the same flash rate. We found that there is not a large amount

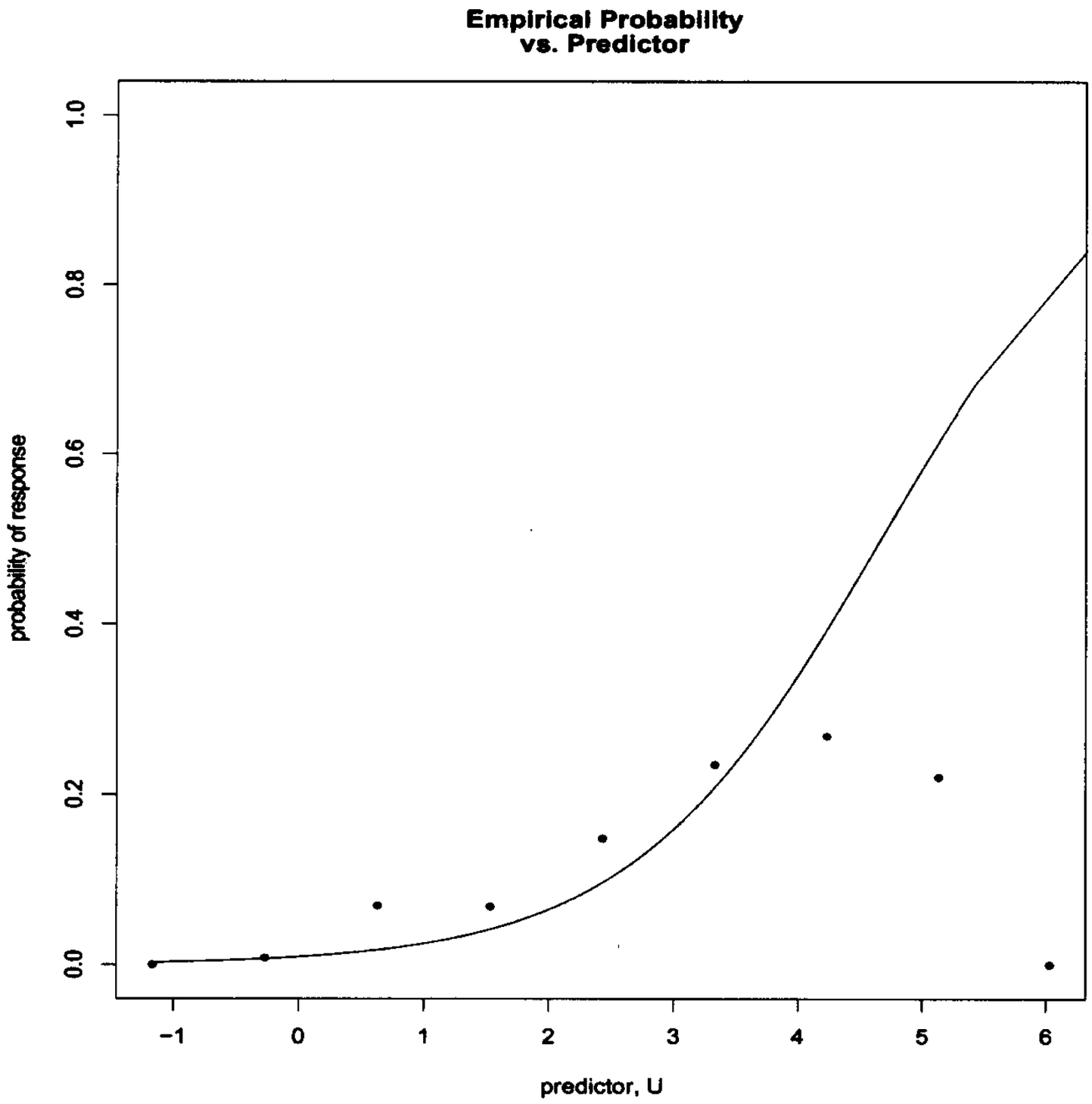


Figure 3.14: Diagnostic plot for a random effects threshold model with logit link fit to 12 runs of simple RT data with flash rate $p_A = 1.0$. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

of variation due to different runs. In addition, the results based on simple pooling of the data sets are very similar to those for the random effects model, which indicates that pooling the data is sufficient in the fitting of threshold models to these RT data.

Chapter 4

Parametric and Threshold Models for a Go-No Go RT Experiment

A go-no go reaction time experiment is more closely related to a real-life task than the simple RT experiment. For a simple RT experiment the sole activity is to decide whether or not there is a flash, and then immediately press a button. In a go-no go reaction time experiment, two types of flashes, black and white, are presented against a grey background, to an observer. Upon perception of a black flash the subject immediately presses a button, while white flashes are ignored by the observer. Thus, a go-no go RT experiment involves the extra activity of deciding when one of two stimulus types occurs.

The observer may correctly press the button in response to a black flash, not respond at all to a black flash, or may press it as as a result of committing one of two possible errors. First, the incorrect response may be due to misperceiving a white flash. Another possibility is pressing the button when neither flash type is presented. This type of error is due to internal noise within the eye-brain-hand system that causes the observer to mistakenly perceive a black stimulus.

The goal of this chapter is two-fold: to present a parametric model and a threshold model for go-no go RT data. First, we introduce a parametric model, and use it in conjunction with nonparametric estimation, to make inferences about the eye-brain-hand system. The second model that we will introduce is a threshold model for go-no

go RT data.

In Section 4.1 we introduce a parametric model for a go-no go RT experiment. We provide an outline of how to simulate data from this model in Section 4.2. In Section 4.3 we describe some point process intensity functions for this model. In the subsequent section we provide the derivations of the intensities. In Section 4.5 we discuss nonparametric intensity estimates and bandwidth selection. We develop parameter estimation methods in Section 4.6, and in Section 4.7 we illustrate our estimation methods on simulated data. Finally, in Section 4.8, we fit our model to ten data sets from go-no go RT experiments with different flash rates. We introduce the threshold model in Section 4.9. Within this section we discuss the model details, statistical methods, diagnostics, and application to data simulated from our parametric model, and to real data. A discussion of our two models concludes this chapter.

4.1 A Parametric Model for the Go-No Go RT Experiment

In a go-no go reaction time experimental run, black and white flashes are presented to a subject as independent homogeneous Poisson processes with rates p_{S_B} and p_{S_W} (per second), respectively. Thus, the entire flash process, which is the superposition of the black and white flash processes, has rate $p_{S_B} + p_{S_W}$.

We denote the number of white flashes in the time interval X by $S_W(X)$, and the set of times of the white flashes is denoted by $S_W = \{S_{W,1}, S_{W,2}, \ldots\}$. Analogous notation is used for the black flash times, S_B, as well as the subject's response times, R. Note that S_B and S_W are taken to be independent stationary Poisson processes.

We use a superposition operation to account for noise internal to the nervous system. We consider the possibility of an observer pressing a button in the absence of a flash. Associated with the black flashes is a noise process, modelled as an independent Poisson process S_N with rate p_N. The noise process is superposed onto the black flash process, so that the superposed process is $S'_B = S_B \bigcup S_N$. Because of the nature of

the experiment, we do not consider a noise process associated with the white flashes.

A thinning operation is used to model a failure to perceive a flash at S'_B or S_W. Each flash is deleted with probability p independently of the other flashes. The deletion probability for the thinning of white flashes is given by

$$P(S_{W,j} \text{ deleted}) = p.$$

For the thinning of the black flashes, in addition to completely random deletions, there may also be non-linear inhibition resulting when two black flashes occur within d time units of each other. The deletion probability for the thinning of black flashes is given by

$$P(S'_{B,j} \text{ deleted}) = \begin{cases} 1, & S'_{B,j+1} - S'_{B,j} < d \\ p, & \text{otherwise} \end{cases}, \tag{4.1.1}$$

so that when a black flash is closely followed by another black flash the two flashes are integrated and observed as one higher intensity flash. Since the white flashes are supposed to be ignored by the observer, analogous interactions between the white flashes are not modelled. We denote the thinned white and black flash processes by S'_W and S''_B, respectively.

Each response is of the form $R_i = S''_i + V_i$, where $\{V_i\}$ is a sequence of independent and identically distributed random variables with common density $f_V(v)$. The random variables V_i correspond to the amounts of time required for the brain and hand to respond to the flash; these are the particular reaction times for each of the events.

The responses are related to the flash processes as follows, where q is the probability of responding to a white flash:

$$S''_{B,j} + V_{0,j} \;\rightarrow\; R_i$$

$$S'_{W,j} + V_{1,j} \;\rightarrow\; R_k, \text{ w.p. } q,$$

where we make the simplifying assumption that the delays $V_{0,j}$ and $V_{1,j}$ are sequences of independent and identically distributed random variables with common density $f_V(v)$; q is an error probability. The reaction times to both kinds of flashes are

assumed to be independent and identically distributed, and the parametric intensity functions that we derive in Section 4.4 depend on this assumption.

We now set out all of the assumptions that we make in this chapter:

Assumptions

1. Reaction times resulting from black (white) flashes are independent random variables.
2. Reaction times from black flashes are independent of those from white flashes.
3. Reaction times from white flashes have the same distribution as those from black flashes.
4. S_B and S_W are independent stationary Poisson processes.
5. All reaction times have a $\mathrm{N}(\mu, \sigma)$ distribution. However, the derivations and estimation methods are also valid for alternative distributions.

4.2 Simulation of Data from the Model

In our simulation of the flash processes, we first simulate the total flash process S, which is a Poisson process with rate $p_{S_B} + p_{S_W}$. Assignment of flash types to each flash time S_i is done by noting that because the black and white flash processes are independent of each other, the probability that a given flash is black is $\frac{p_{S_B}}{p_{S_B}+p_{S_W}}$. Likewise, the probability that a given flash is white is $\frac{p_{S_W}}{p_{S_B}+p_{S_W}}$. In our algorithm, ID_S denotes a stimulus identification vector, where $ID_{S,i}$ identifies the stimulus type for S_i. We assign $ID_{S,i} \leftarrow 0$ when the flash at time S_i is black, and $ID_{S,i} \leftarrow 1$ when the flash at time S_i is white.

In the case that $p_N > 0$, letting T be the time length of the experiment, the number of events from the noise process is a Poisson random variable with mean $p_N T$. We denote the number of events in the noise process by N_N. The noise process is simulated by generating N_N uniform random variables on the interval $(0, T)$. Each noise event time is assigned a black identification label of '0'.

Both the noise event times and corresponding identification labels are appended to the flash times and identification labels. The stimulus times (noise and flashes) and identification labels are sorted in ascending order by stimulus times. This is done by sorting the vector of stimulus times, and changing the positions of the identifications for each change in position of the corresponding stimulus times.

In Algorithm 4.2.1 the vector R consists of the response times, and we denote the number of responses by N_R.

The data set will be a list consisting of three vectors, not necessarily the same length, consisting of stimulus times and response times, as well as labels to indicate the type of stimulus. That is, each black stimulus time has a label of '0', while a label of '1' identifies those that are white.

Algorithm 4.2.1. *The following steps can be used to generate go-no go RT data consisting of N_S flashes with black flash rate p_{S_B}, white flash rate p_{S_W}, noise rate p_N, thinning parameter d, thinning probability p, and $N(\mu, \sigma)$ distributed reaction times.*

1. *Generate stimulus sequences consisting of a total of N_S black and white stimuli.*

   (a) *Generate N_S random exponential variates with mean $\frac{1}{p_{S_B}+p_{S_W}}$.*

   (b) *Compute the N_S cumulative sums of the exponential variates. Assign these times to S.*

   (c) *Set*

   $$p_0 \leftarrow \frac{p_{S_B}}{p_{S_B} + p_{S_W}}.$$

   (d) *Assign the stimulus types. For $i = 1, 2, \ldots, N_S$, set*

   $$ID_S[i] \leftarrow \begin{cases} 0 & \text{with probability } p_0 \\ 1 & \text{with probability } 1 - p_0 \end{cases}$$

   (e) *Set $S_B \leftarrow \{S[i] : ID_S[i] = 0\}$ and $S_W \leftarrow \{S[i] : ID_S[i] = 1\}$.*

2. *Add noise. If $p_N > 0$, then*

   (a) *Generate a Poisson $p_N T$ random variate and assign to N_N.*

(b) *If* $N_N = 0$, *go to Step 3.*

(c) *Assign* $N_S \leftarrow N_S + N_N$.

(d) *Generate* N_N *random uniform variates on* $[0, T]$. *Assign these values to* N.

(e) *Assign the black stimulus type to the noise. For* $i = 1, 2, \ldots, N_N$, *set*

$$ID_S[i] \leftarrow 0.$$

(f) *Set* $S' \leftarrow S \cup N$ *and* $ID_{S'_N} \leftarrow ID_S \cup ID_{S_N}$.

(g) *Sort* S' *and* $ID_{S'_N}$ *in ascending order of* S'.

3. *Thinning and translating.*

(a) *Set* $N_R \leftarrow 0$.

(b) *For* $i = 2, \ldots, N_{S_B}$, *if* $(S'_B[i] - S'_B[i-1] > d)$, *then do the following with probability* $1 - p$:

i. *Assign* $N_R \leftarrow N_R + 1$.

ii. *Generate a* $N(\mu, \sigma^2)$ *random variable* V.

iii. *Assign* $R[N_R] \leftarrow S'_B[i-1] + V$

(c) *For* $i = 2, \ldots, N_{S_W}$, *do the following with probability* $(1-p)q$:

i. *Assign* $N_R \leftarrow N_R + 1$.

ii. *Generate a* $N(\mu, \sigma^2)$ *random variable* V.

iii. *Assign* $R[N_R] \leftarrow S_W[i-1] + V$

4. *Sort* R *in ascending order.*

5. *Return* S, ID_S, *and* R.

We used the programming language Fortran (Lahey Computer Systems, Inc. (1999)) to implement our algorithm.

4.3 Point Process Intensity Functions

The response rate, or first order intensity of the R process is denoted p_R. We denote the response-response second order intensity by $p_{RR}(u)$.

There are two stimulus-response second order intensities corresponding to each flash type with the response process, which we denote $p_{S_BR}(u)$ and $p_{S_WR}(u)$. The intensity $p_{S_BR}(u)$ corresponds to the intensity of a response following a black flash at a lag of u time units, while $p_{S_WR}(u)$ coincides with the intensity of a white flash followed by a response.

Under our model, we show that the Brillinger mixing condition holds by using an argument similar to that given in Section 3.1.1. By the same reasoning as in Section 3.1.1, which applies a result of Heinrich (1988), we have that each of the three processes S_B, S_W, and R is Brillinger mixing.

In order to show that the point process (S_B, S_W, R) is Brillinger mixing, we consider the point process $S_B \bigcup S_W \bigcup R$, which we denote by C. The process C is a Poisson cluster process with cluster centers at $S_i \in S_B \bigcup S_W$. When there is no thinning in our model (i.e. $d = 0$ and $p = 0$), each cluster center S_i generates two points: $\{S_i + 0, S_i + V_i\}$. Note that $S_i + V_i \in R$. In the case that there is thinning, there is also the possibility that S_i generates only one point: $\{S_i + 0\}$. Therefore, regardless of the presence or absence of thinning, $N_s(X)$ takes on values in $\{1, 2\}$, and it follows that $E[N_s^k(X)] \leq 2^k < \infty$. Thus, $S_B \bigcup S_W \bigcup R$ is Brillinger mixing, from which we conclude that the three vector valued point process (S_B, S_W, R) is also Brillinger mixing.

For our parametric model the response rate is given by

$$p_R = (1-p)[qp_{S_W} + (p_{S_B} + p_N)e^{-d(p_{S_B}+p_N)}]. \tag{4.3.1}$$

The second order intensities are given by:

$$\begin{aligned} p_{S_B R}(u) &= (1-p)p_{S_B}[qp_{S_W} \\ &\quad +e^{-d(p_{S_B}+p_N)}\{(p_{S_B}+p_N)P(V \notin (u,u+d)) + f(u)\}] \qquad (4.3.2) \\ p_{S_W R}(u) &= (1-p)p_{S_W}[(p_{S_B}+p_N)e^{-d(p_{S_B}+p_N)} + q(p_{S_W}+f(u))] \qquad (4.3.3) \\ p_{RR}(u) &= (1-p)^2[P(|V_2-V_1-u|>d)\{(p_{S_B}+p_N)^2 e^{-2d(p_{S_B}+p_N)}\} \\ &\quad +q^2 p_{S_W}^2 + 2q(p_{S_B}+p_N)p_{S_W}e^{-d(p_{S_B}+p_N)}], \qquad (4.3.4) \end{aligned}$$

where V, V_1, and V_2 are independent random variables having common density $f_V(v)$.

In the special case of $q = 0$, so that white flashes are never responded to, the expressions simplify to:

$$\begin{aligned} p_R &= (1-p)(p_{S_B}+p_N)e^{-d(p_{S_B}+p_N)} \\ p_{S_B R}(u) &= p_{S_B}p_R\left\{P(V \notin (u,u+d)) + \frac{f(u)}{p_{S_B}+p_N}\right\} \\ p_{S_W R}(u) &= p_{S_W}p_R \\ p_{RR}(u) &= p_R^2 P(|V_2-V_1-u|>d). \end{aligned}$$

The expressions for p_R, $p_{S_B R}(u)$, and $p_{RR}(u)$ are identical to those for the simple RT model introduced by Braun et al. (2003). In addition, since in this case the response and white flash processes are independent of each other, as expected, $p_{S_W R}(u) = p_{S_W}p_R$, from which it can be seen that $q_{S_W R}(u) = 0$ for all u.

Graphs of the two stimulus-response second order intensities with three values of q at two values of d are provided in Figure 4.1. These intensities have a peak at the mode of $f(u)$, which increases in height as q increases, with the effect most distinct for $p_{S_W R}(u)$. As q approaches zero the intensity $p_{S_W R}(u)$ flattens and is constant for $q = 0$. Thus, information about the error probability q is contained in $p_{S_W R}(u)$. As the d parameter increases the curve to the left of the peak of $p_{S_B R}(u)$ is depressed. Note that the d parameter has the same effect on the stimulus-response second order intensity for the simple RT model described in Braun et al. (2003).

Figure 4.2 displays plots of the second order response-response intensity for three values of d at two values of q. Regardless of the value of q, when $d = 0$ the intensities

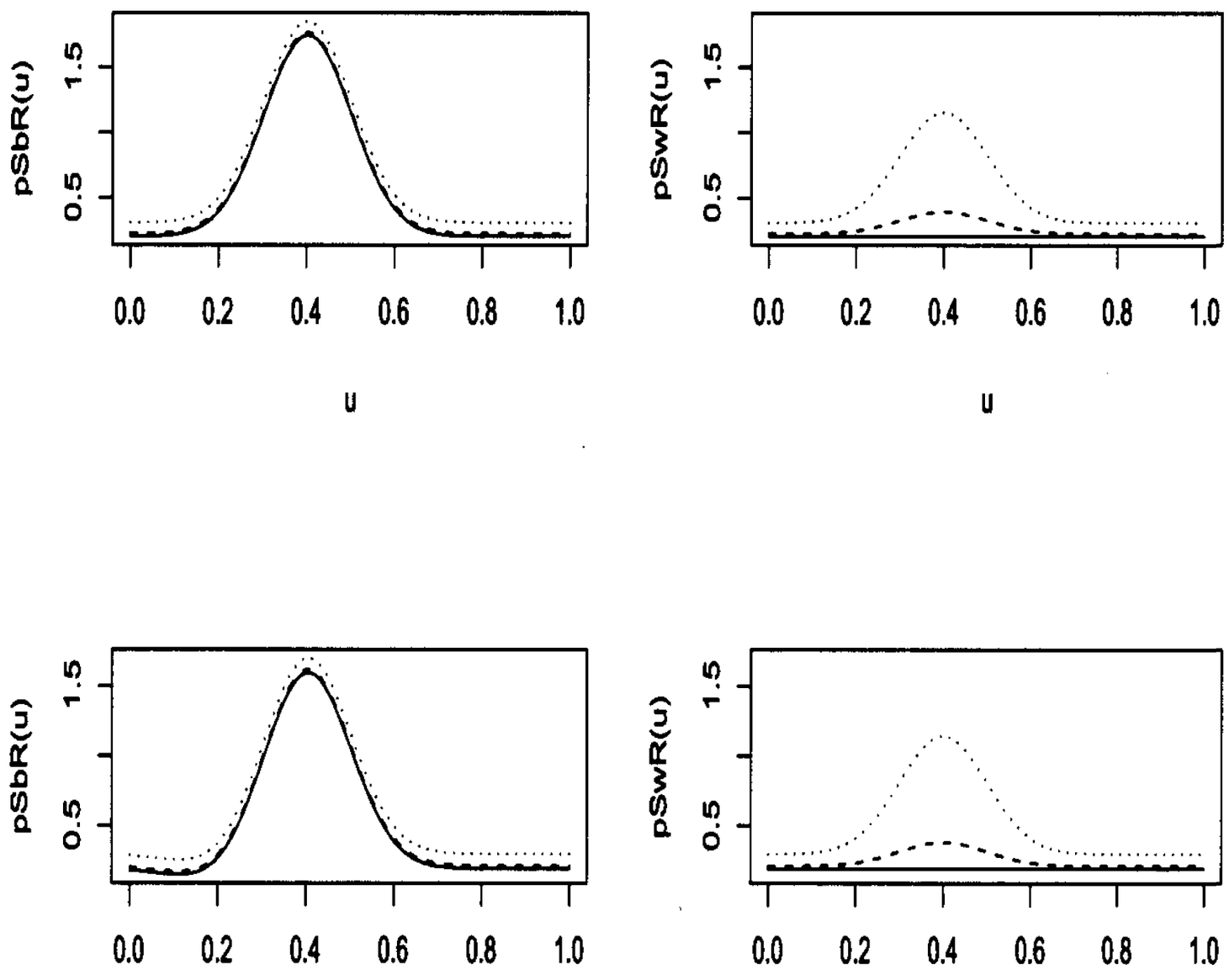


Figure 4.1: The second-order stimulus-response intensity functions with $p_{S_W} = .5$, $p_{S_B} = .5$, μ=.4, σ=.1, p=.15, $p_N = 0$, $q = 0$ (solid line), $q = .1$ (dashed line), and $q = .5$ (dotted line), and $d = .1$ (top panel), $d = .25$ (bottom panel)

are constant. Otherwise, there is a trough centered at zero, which widens as d increases. The effect of q is not as noticeable, but as q increases the curve slowly shifts upwards and the depth of the trough decreases slightly. Thus, the response-response intensity contains information on the nonlinear inhibition parameter d.

4.4 Derivations of the Intensity Functions

Before proceeding to the derivations we briefly review the notation for our point processes:

- S_B: black flash process; a Poisson process with rate p_{S_B}

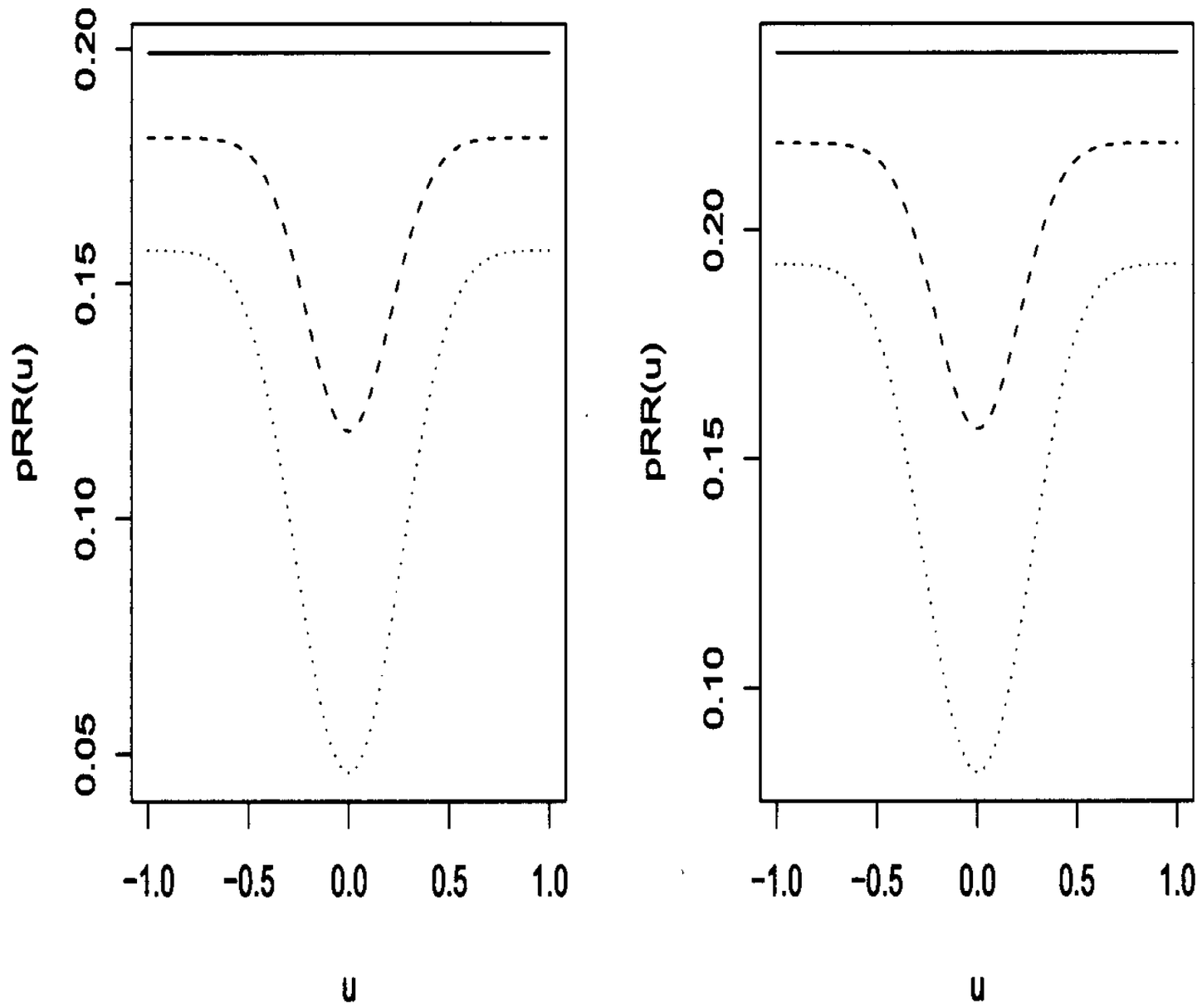


Figure 4.2: The second-order response-response intensity function in the normal case with $p_{S_W} = .5$, $p_{S_B} = .5$, μ=.4, σ=.1, p=.15, $p_N = 0$, $d = 0$ (solid line), $d = .1$ (dashed line), and $d = .25$ (dotted line), and $q = .05$ (left panel), $q = .15$ (right panel)

- S_N: noise process for black flashes; a Poisson process with rate p_N
- S'_B: $S_B \bigcup S_N$, black flash process superposed with noise process; a Poisson process with rate $p_{S_B} + p_N$
- S''_B: the resulting process after thinning of S'_B; each event contributes a response
- S_W: white flash process; a Poisson process with rate p_{S_W}
- S'_W: the resulting process after thinning of S_W; each event contributes a response with error probability q
- R: response process consisting of both correct and error responses.

4.4.1 First-Order Response Intensity

We first derive the expression for the response rate. We have

$$p_R = \lim_{h \to 0} h^{-1} E[R(u, u+h)]. \tag{4.4.1}$$

As a result of the translation mechanism in the model, we have

$$R(u, u+h) = \sum_i 1_{\{S''_{B,i} \in (u-V_i, u-V_i+h)\}} + \sum_i 1_{\{S'_{W,i} \in (u-V_i, u-V_i+h)\}} A_i,$$

where

$$A_i = 1_{\{\text{response|white flash at} S'_i\}},$$

$S'_i = (S'''_B \bigcup S'_W)_i$ and the V_i represent independent random variables from a population having density $f(v)$. The probability of a response given a white flash is the error probability q. Thus,

$$E[R(u, u+h)] = E\left[\sum_i 1_{\{S''_{B,i} \in (u-V_i, u-V_i+h)\}}\right] + qE\left[\sum_i 1_{\{S'_{W,i} \in (u-V_i, u-V_i+h)\}}\right]. \tag{4.4.2}$$

The first term in (4.4.2) corresponds to the response process rate for correct responses, while the second term coincides with the rate of the error response process.

White flashes are deleted with probability p independently of the other flashes, so we have

$$E[S'_W(dx_1)] = E[S_W(dx_1) 1_{\{U_{y_1 - v} > p\}}] = (1-p) p_{S_W}\, dx_1,$$

where $U_{y_1 - v}$ is an independent uniform random variable that is independent of the flash and response processes. Thus the thinned white flashes are a Poisson process with rate $p_{S'_W} = p_{S_W}(1-p)$.

A black event is deleted with probability p when another black event occurs within d time units after it, so we have

$$\begin{aligned} E[S''_B(d(x_1))] &= E[S'_B(d(x_1)) 1_{\{U_{x_1} > p\}} 1_{\{S'_B(x_1, x_1+d)=0\}}] \\ &= (1-p)(p_{S_B} + p_N) e^{-d(p_{S_B} + p_N)}\, dx_1. \end{aligned} \tag{4.4.3}$$

Thus, we have

$$
\begin{aligned}
&E[R(u,u+h)] \\
&= qE[\sum_i 1_{\{S'_{W,i}\in(u-V_i,u-V_i+h)\}}] + E[\sum_i 1_{\{S''_{B,i}\in(u-V_i,u-V_i+h)\}}] \\
&= qE[E[\sum_i 1_{\{S'_{W,i}\in(u-V_i,u-V_i+h)\}}|S'_W]] + E[E[\sum_i 1_{\{S''_{B,i}\in(u-V_i,u-V_i+h)\}}|S''_B]] \\
&= qE[E[E[\sum_i 1_{\{S'_{W,i}\in(u-v,u-v+h)\}}|V,S'_W]] + E[E[E[\sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}|V,S''_B]]] \\
&= qE[E[\sum_i \int 1_{\{S'_{W,i}\in(u-v,u-v+h)\}} f(v)\,dv|S'_W]] \\
&\quad + E[E[\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}} f(v)\,dv|S''_B]] \\
&= qE[\sum_i \int 1_{\{S'_{W,i}\in(u-v,u-v+h)\}} f(v)\,dv] \\
&\quad + E[\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}} f(v)\,dv] \qquad (4.4.4) \\
&= qE[\int \sum_i 1_{\{S'_{W,i}\in(u-v,u-v+h)\}} f(v)\,dv] \\
&\quad + E[\int \sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}} f(v)\,dv] \qquad (4.4.5) \\
&= qE[\int S'_W(u-v,u-v+h) f(v)\,dv] \\
&\quad + E[\int S''_B(u-v,u-v+h) f(v)\,dv] \qquad (4.4.6)
\end{aligned}
$$

where (4.4.4) is obtained by using the fact that the V_i are i.i.d. with density $f(\cdot)$. By application of the Tonelli-Fubini theorem (see Jacod and Protter (1991), pp. 63-64), the integral and sum are reversed to get (4.4.5).

Since S'_W, and S''_B are counting measures, and for a counting measure $A(\cdot)$ we have $A(0,h) = \int_0^h A(dx_1)$ (see Daley and Vere-Jones (2003)), (4.4.6) may be expressed as

$$qE[\int \int_u^{u+h} S'_W(d(y_1-v)) f(v)\,dv] + E[\int \int_u^{u+h} S''_B(d(y_1-v)) f(v)\,dv].$$

By noting that the integrand is non-negative the order of integration and expectation

may be reversed yielding

$$\begin{aligned} & E[R(u,u+h)] \\ &= q\int_u^{u+h}\int E[S'_W(d(y_1-v))]f(v)\,dv + \int_u^{u+h}\int E[S''_B(d(y_1-v))]f(v)\,dv. \end{aligned}$$

Using (4.4.3) and the stationarity of S''_B and S'_W it follows that

$$\begin{aligned} & E[R(u,u+h)] \\ &= q\int_u^{u+h}\int (1-p)p_{S_W}\,dy_1 f(v)\,dv \\ &\quad + \int_u^{u+h}\int (1-p)(p_{S_B}+p_N)e^{-d(p_{S_B}+p_N)}\,dy_1 f(v)\,dv. \\ &= h(1-p)[qp_{S_W}+)(p_{S_B}+p_N)e^{-d(p_{S_B}+p_N)}]\int f(v)\,dv. \end{aligned} \tag{4.4.7}$$

Substitution of (4.4.7) into (4.4.1), and using the fact that $f(v)$ is a probability density function, we see that the first order intensity for the R process is given by

$$p_R = (1-p)[qp_{S_W} + (p_{S_B}+p_N)e^{-d(p_{S_B}+p_N)}].$$

4.4.2 Stimulus-Response Second-Order Intensities

We now derive expression (4.3.2) for $p_{S_B R}(u)$. By a similar argument expression (4.3.3) for $p_{S_W R}(u)$ is derived, so for this case only the key steps are outlined.

The intensity $p_{S_B R}(u)$ is generally a function of two variables, but it reduces to a function of one variable because of stationarity. That is, letting $u = y - x$,

$$p_{S_B R}(x,y) = \lim_{h\to 0} h^{-2}E[S_B(x,x+h)R(y,y+h)]$$

may be expressed as

$$p_{S_B R}(u) = \lim_{h\to 0} h^{-2}E[S_B(0,h)R(u,u+h)]. \tag{4.4.8}$$

In the following we assume that the reaction times V_i from the white and black flashes are independently and identically distributed with the same distribution $f(v)$.

That is, this result depends on Assumptions 1, 2, and 3.

$$
\begin{aligned}
& E[S_B(0,h)R(u,u+h)] \\
= \; & E[S_B(0,h)(qE[\sum_i 1_{\{S'_{W,i}\in(u-V_i,u-V_i+h)\}}|S'_W,S''_B] \\
& +E[\sum_i 1_{\{S''_{B,i}\in(u-V_i,u-V_i+h)\}}|S'_W,S''_B])] \\
= \; & E[S_B(0,h)(qE[E[\sum_i 1_{\{S'_{W,i}\in(u-v,u-v+h)\}}|V,S'_W]] \\
& +E[E[\sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}|V,S''_B]])] \qquad (4.4.9) \\
= \; & E[S_B(0,h)(qE[\sum_i \int 1_{\{S'_{W,i}\in(u-v,u-v+h)\}}f(v)\,dv|S'_W] \\
& +E[\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}f(v)\,dv|S''_B])] \\
= \; & E[S_B(0,h)(q\sum_i \int 1_{\{S'_{W,i}\in(u-v,u-v+h)\}}f(v)\,dv \\
& +\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}f(v)\,dv)] \\
= \; & E[S_B(0,h)(q\int \sum_i 1_{\{S'_{W,i}\in(u-v,u-v+h)\}}f(v)\,dv \\
& +\int \sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}f(v)\,dv)] \qquad (4.4.10) \\
= \; & E[S_B(0,h)(q\int S'_W(u-v,u-v+h)f(v)\,dv \\
& +\int S''_B(u-v,u-v+h)f(v)\,dv)] \qquad (4.4.11)
\end{aligned}
$$

where (4.4.9) is obtained by using the fact that the V_i are i.i.d. with density $f(\cdot)$. By application of the Tonelli-Fubini theorem, the integral and sum are reversed to get (4.4.10).

Since S_W, S_B, S'_W, and S''_B are counting measures (4.4.11) may be expressed as

$$
\begin{aligned}
& qE[\int_0^h S_B(dx_1)\int\int_u^{u+h} S'_W(d(y_1-v))f(v)\,dv] \\
+ \; & E[\int_0^h S_B(dx_1)\int\int_u^{u+h} S''_B(d(y_1-v))f(v)\,dv].
\end{aligned}
$$

By noting that the integrand is non-negative the order of integration and expectation

may be reversed yielding

$$\begin{aligned} & E[S_B(0,h)R(u,u+h)] \\ &= q\int_u^{u+h}\int_0^h\int E[S_B(dx_1)S'_W(d(y_1-v))]f(v)\,dv \\ &\quad + \int_u^{u+h}\int_0^h\int E[S_B(dx_1)S''_B(d(y_1-v))]f(v)\,dv. \end{aligned} \tag{4.4.12}$$

Because of the thinning assumption, the two expectations in (4.4.12) can be expressed as

$$E[S_B(dx_1)S'_W(d(y_1-v))] = E[S_B(dx_1)S_W(d(y_1-v))1_{\{U_{y_1-v}>p\}}], \tag{4.4.13}$$

and

$$\begin{aligned} & E[S_B(dx_1)S''_B(d(y_1-v))] \\ &= E[S_B(dx_1)S'_B(d(y_1-v))1_{\{U_{y_1-v}>p\}}1_{\{S'_B(y_1-v,y_1-v+d)=0\}}]. \end{aligned} \tag{4.4.14}$$

Using the independence of U_{y_1-v} and the fact that S_B and S_W are independent Poisson processes with rates p_{S_B} and p_{S_W}, respectively, (4.4.13) can be re-written as

$$E[S_B(dx_1)S'_W(d(y_1-v))] = (1-p)p_{S_W}p_{S_B}\,dx_1\,dy_1. \tag{4.4.15}$$

To simplify (4.4.14), we note that if $0 < x_1-(y_1-v) < d$, then $S'_B(y_1-v)$ is deleted, because there is then a point at least d time units after $S'_B(y_1-v)$. Thus, we get the expression

$$\begin{aligned} & E[S_B(dx_1)S''_B(d(y_1-v))] \\ &= (1-p)e^{-d(p_{S_B}+p_N)}(C_0(x_1,y_1,v)+C_1(x_1,y_1,v))\,dx_1\,dy_1, \end{aligned} \tag{4.4.16}$$

where

$$\begin{aligned} C_0(x_1,y_1,v) &= p_{S_B}(p_{S_B}+p_N)1_{\{x_1\notin(y_1-v,y_1-v+d)\}}, \\ C_1(x_1,y_1,v) &= p_{S_B}\delta(y_1-v-x_1). \end{aligned}$$

The terms C_0 and C_1 account for the events that the condition $0 < x_1-(y_1-v) < d$ fails and holds, respectively. If the condition holds, then (4.4.16) is nonzero only when $x_1 = y_1 - v$, which is identified by the Dirac delta in C_1.

Substitution of (4.4.15) and (4.4.16) into (4.4.12) gives

$$\begin{aligned} & E[S_B(0,h)R(u,u+h)] \\ &= (1-p)\Big(q\int_0^h\int_u^{u+h}\int p_{S_B}p_{S_W}f(v)\,dv\,dy_1\,dx_1 \\ &+ \int_0^h\int_u^{u+h}\int e^{-d(p_{S_B}+p_N)}[C_0(x_1,y_1,v)+C_1(x_1,y_1,v)]f(v)\,dv\,dy_1\,dx_1\Big) \quad (4.4.17) \end{aligned}$$

Using the property of the Dirac delta with $x = v$ and $x_0 = y_1 - x_1$, and noting that since $f(v)$ is a density function we have $\int f(v)\,dv = 1$, (4.4.17) simplifies to

$$\begin{aligned} E[S_B(0,h)R(u,u+h)] &= (1-p)p_{S_B}\Big(qh^2p_{S_W} \\ &+e^{-d(p_{S_B}+p_N)}\int_0^h\int_u^{u+h}\Big[(p_{S_B}+p_N)\int 1_{\{v\notin(y_1-x_1,y_1-x_1+d)\}}f(v)\,dv \\ &+f(y_1-x_1)\Big]\,dy_1\,dx_1\Big) \quad (4.4.18) \end{aligned}$$

Finally, substituting (4.4.18), after two applications of the mean value theorem for integrals to the double integral involving $f(y_1 - x_1)$, we get an expression for $p_{S_BR}(u)$:

$$\begin{aligned} p_{S_BR}(u) &= \lim_{h\to 0} E[S_B(0,h)R(u,u+h)] \\ &= (1-p)p_{S_B}[qp_{S_W} + e^{-d(p_{S_B}+p_N)}\{(p_{S_B}+p_N)P(V\notin(u,u+d)) + f(u)\}]. \end{aligned}$$

In a similar manner $p_{S_WR}(u)$ can be obtained, by noting that by similar arguments we have:

$$\begin{aligned} E[S_W(dx_1)S'_W(d(y_1-v))] &= E[S_W(dx_1)S_W(d(y_1-v))1_{\{U_{y_1-v}>p\}}] \\ &= (1-p)[p^2_{S_W} + p_{S_W}\delta(x_1-y_1+v)], \end{aligned}$$

and

$$\begin{aligned} E[S_W(dx_1)S''_B(d(y_1-v))] &= E[S_W(dx_1)S'_B(d(y_1-v))1_{\{U_{y_1-v}>p\}}1_{\{S'_B(y_1-v,y_1-v+d)=0\}}] \\ &= p_{S_W}(p_{S_B}+p_N)(1-p)e^{-d(p_{S_B}+p_N)}, \end{aligned}$$

so that

$$p_{S_WR}(u) = (1-p)p_{S_W}[q\{p_{S_W}+f(u)\} + (p_{S_B}+p_N)e^{-d(p_{S_B}+p_N)}].$$

4.4.3 Response-Response Second Order Intensities

Arguments similar to those used in the derivation of the second-order stimulus-response second-order intensities are used in the derivation of the response auto-intensity function $p_{RR}(u)$.

$$
\begin{aligned}
& E[R(x,x+h)R(y,y+h)] \\
= \; & E[\{q\int\int_x^{x+h} S'_W(d(x_1-v_1))f(v_1)\,dv_1 + \int\int_x^{x+h} S''_B(d(x_1-v_1))f(v_1)\,dv_1\} \\
& \times\{q\int\int_y^{y+h} S'_W(d(y_1-v_2))f(v_2)\,dv_2 + \int\int_y^{y+h} S''_B(d(y_1-v_2))f(v_2)\,dv_2\}] \\
= \; & (1-p)^2\Big[q^2\int\int\int_x^{x+h}\int_{y:y_1\neq x_1}^{y+h} E[S_W(d(x_1-v_1))S_W(d(y_1-v_2))] \\
& \times f(v_1)f(v_2)\,dx_1\,dy_1\,dv_1\,dv_2 + p_R|(x,x+h)\bigcap(y,y+h)|_L \\
+ \; & \int\int\int_x^{x+h}\int_{y:y_1\neq x_1}^{y+h} E[S'_B(d(x_1-v_1))S'_B(d(y_1-v_2))1_{\{S'_B(x_1-v_1,x_1-v_1+d)=0\}} \\
& \times 1_{\{S'_B(y_1-v_2,y_1-v_2+d)=0\}}]f(v_1)f(v_2)\,dx_1\,dy_1\,dv_1\,dv_2 + p_R|(x,x+h)\bigcap(y,y+h)|_L \\
+ \; & 2q\int\int\int_x^{x+h}\int_y^{y+h} E[S_W(d(x_1-v_1))S'_B(d(y_1-v_2)) \\
& \times 1_{\{S'_B(y_1-v_2,y_1-v_2+d)=0\}}]f(v_1)f(v_2)\,dx_1\,dy_1\,dv_1\,dv_2\Big]
\end{aligned}
$$

Making use of the fact that S'_B and S_W are independent Poisson processes with respective rates $p_{S_B}+p_N$ and p_{S_W} allows us to write the second term in the sum as

$$\int\int h^2(p_{S_B}+p_N)^2 e^{-2d(p_{S_B}+p_N)} 1_{\{x-v_1\notin(y-v_2,y-v_2+d)\}} 1_{\{y-v_2\notin(x-v_1,x-v_1+d)\}}]f(v_1)f(v_2)\,dv_1\,dv_2$$

After division by h^2 and letting $h\to 0$, we obtain an expression for $p_{RR}(u)$, where $u=y-x$:

$$
\begin{aligned}
p_{RR}(u) \; = \; & (1-p)^2[q^2p_{S_W}^2 + (p_{S_B}+p_N)^2 e^{-2d(p_{S_B}+p_N)}P(|V_2-V_1-u|>d) \\
& +2qp_{S_W}(p_{S_B}+p_N)e^{-d(p_{S_B}+p_N)}].
\end{aligned}
$$

4.5 Nonparametric Intensity Estimation

As for simple RT, the first-order response intensity can be estimated by

$$\hat{p}_R = R([0,T])/T, \tag{4.5.1}$$

where T is the time of the last response ($T = \max(R_k)$). Let $K(x)$ be a symmetric second order smoothing kernel with compact support, and h be an appropriately chosen bandwidth. Then the second-order nonparametric intensity estimates are given by

$$\hat{p}_{SR}(u) = \frac{1}{Th} \sum_i \sum_j K\left(\frac{u - (R_j - S_i)}{h}\right), \tag{4.5.2}$$

where S denotes one of the stimulus processes, and

$$\hat{p}_{RR}(u) = \frac{1}{Th} \sum_i \sum_{j \neq i} K\left(\frac{u - (R_j - R_i)}{h}\right). \tag{4.5.3}$$

In this analysis we use the biweight kernel, $K(x) = (15/16)(1-x^2)^2 1_{\{|x|<1\}}$. Brillinger (1975b) has shown that under cumulant mixing conditions, the intensity function estimates are asymptotically normal as $h \to 0$ and $Th \to \infty$, with mean

$$E[\hat{p}_{SR}(u)] = \int K(u_1) p_{SR}(u - u_1 h)\, du_1 \tag{4.5.4}$$

and asymptotic variance

$$(Th)^{-1} \left(\int K^2(u_1)\, du_1 \right)^2 p_{SR}(u) \tag{4.5.5}$$

The bias is of order $O(h^2)$.

We assume that the cumulant mixing condition holds, so that the above holds for each estimate, replacing p_{SR} by the desired second-order intensity. This is justifiable since the condition holds under the model (see Section 4.3), and the model is a rough approximation to the data.

4.5.1 Bandwidth Selection

We find optimal bandwidths for the intensities by minimizing the asymptotic mean integrated squared error (AMISE) with respect to h, the bandwidth. A general expression of the AMISE optimal bandwidth for a second-order intensity is derived in Section 2.6.1, and in the case of the biweight kernel we have

$$h_{AMISE}(p_{SR}) = \left[\frac{35 \int p_{SR}(u)\, du}{T \int (\frac{\partial^2}{\partial^2 u} p_{SR}(u))^2\, du} \right]^{1/5}. \tag{4.5.6}$$

Using our fitted model, we can obtain estimates of $h_{AMISE}(p_{SR})$, which are different for each intensity. Based on the expression for $p_{S_BR}(u)$, given in (4.3.2), it can be seen that dependency on u is through

$$w_B(u) = P(V \notin (u, u+d)) + \frac{f(u)}{p_{S_B} + p_N},$$

which, under the assumption of a normal delay distribution, has second derivative

$$w''_B(u) = \frac{u+d-\mu}{\sigma^2} f(u+d) + \left[\frac{1}{(p_{S_B} + p_N)\sigma^2} \left(\frac{(u-\mu)^2}{\sigma^2} - 1 \right) - \frac{(u-\mu)}{\sigma^2} \right] f(u),$$

where $f(\cdot)$ is the Normal density with mean μ and variance σ^2. Similarly, from (4.3.3) it can be seen that $p_{S_WR}(u)$ is dependent on u only through

$$w_W(u) = \frac{f(u)}{p_{S_W}},$$

for which we have

$$w''_W(u) = \frac{1}{p_{S_W}\sigma^2} \left(\frac{(u-\mu)^2}{\sigma^2} - 1 \right) f(u).$$

Thus,

$$\frac{\partial^2}{\partial^2 u} p_{S_BR}(u) = (1-p) p_{S_B} (p_{S_B} + p_N) e^{-d(p_{S_B}+p_N)} w''_B(u),$$

and

$$\frac{\partial^2}{\partial^2 u} p_{S_WR}(u) = (1-p) q p_{S_W}^2 w''_W(u).$$

Similarly, using the expression for $p_{RR}(u)$ given in (4.3.4) it can be seen that this intensity depends on u by the term:

$$s(u) = P(V_2 - V_1 - u| > d),$$

which has second derivative

$$s''(u) = \frac{1}{2\sigma^2} \Big[(u+d)g(u+d) - (u-d)g(u-d) \Big],$$

where $g(\cdot)$ denotes the Normal density with mean 0 and variance $2\sigma^2$. Thus,

$$\frac{\partial^2}{\partial^2 u} p_{RR}(u) = (1-p)^2 (p_{S_B} + p_N)^2 e^{-2d(p_{S_B}+p_N)} s''(u).$$

Finally, for a given second-order intensity, substitution of the intensity expression and corresponding second derivative into (4.5.6) yields the exact form of an optimal bandwidth.

4.6 Parameter Estimation

For our parametric model the likelihood function does not have a closed-form expression, so parameter estimation via maximum likelihood is very complicated, and may not even be possible. Assumptions 1 through 5 allow us to use a quick approximation method that is a modification of the one used by Braun et al. (2003) to estimate μ, σ, and the noise process rate p_N for simple RT data.

In Braun et al. (2003) an iterative algorithm is used to obtain $\hat{\mu}$, $\hat{\sigma}$, and $\hat{p}_N$ for simple RT data. Assuming independent and identically distributed reaction times from both the black and white flashes allows us to consider only the times of the flashes and responses, and not the types of flashes. In their algorithm, estimates of μ and σ are obtained after each iteration. However, the locations of the peaks of the intensities $p_{S_W R}(u)$ and $p_{S_B R}(u)$ correspond to the modes of the white and black reaction times, respectively, which we assume to be equal. Therefore, the locations of the peaks for the corresponding nonparametric estimates should be near the value of the mean reaction times, since we are assuming a symmetric distribution. Based on simulated data we have found more variability in the location of the peak for $p_{S_W R}(u)$ - this is discussed further in Section 4.7. Therefore, we focus on $p_{S_B R}(u)$, and as a modified algorithm, we choose the location of the peak of $\hat{p}_{S_B R}(u)$, which we denote μ_0, as an initial value for the estimation of μ. The estimates of σ and the noise rate p_N are as in the algorithm of Braun et al. (2003). The modified algorithm is given as Algorithm 4.6.1. The algorithm is dependent on the assumption that the reaction times are normally distributed, but it can be adapted to handle other distributions, if necessary.

Algorithm 4.6.1. *For RT data (S, R) the parameters μ, σ, and p_N can be estimated as follows:*

1. *Find the location of the peak in a nonparametric intensity estimate of $p_{S_B R}(u)$ and set μ_0 as the peak location.*

2. *Match each response time R_j with a flash time preceding it, $\hat{S}''_j$, such that $R_j - \hat{S}''_j$ is closest to μ_0, to obtain estimates of the reaction times: $\hat{V}_j = R_j - \hat{S}''_j$.*

3. *Estimate μ and σ by the mean and standard deviation of the $\hat{V}_j$.*

4. *Let $T = R_{N_R}$, where N_R is the number of responses. Estimate the noise rate p_N by:*

$$\hat{p}_N = \frac{|\{\hat{S}''_j : \hat{S}''_j = \hat{S}''_k\}|}{T},$$

where $|\cdot|$ denotes set cardinality.

Examination of plots of the intensities $p_{S_W R}(u)$ and $p_{RR}(u)$ indicate that these functions may be useful in obtaining estimates of q and d, respectively. Furthermore, for long-run experiments (large T values) the bias of each of the nonparametric intensity estimates is near zero, and one expects the difference between the nonparametric and parametric estimates to be minimal. We obtain estimates of d, q and p by minimizing the differences between parametric (based on model expression with parameter estimates) and corresponding nonparametric estimates at a certain point.

Manipulation of (4.3.1), the expression for p_R, allows us to express $\hat{p}$ in terms of $\hat{d}$ and $\hat{q}$ as follows

$$\hat{p} = 1 - \frac{p_R}{(p_{S_B} + \hat{p}_N)e^{-\hat{d}(p_{S_B}+\hat{p}_N)} + \hat{q}p_{S_W}}, \tag{4.6.1}$$

where $\hat{p}_R$ is a nonparametric estimate given by (4.5.1) and $\hat{p}_N$ is the estimate obtained from Algorithm 4.6.1. If (4.6.1) is negative then we set $\hat{p}$ as zero.

Motivated by the effect of q on the behavior of $p_{S_W R}(u)$, manipulation of (4.3.3) with p replaced by (4.6.1) yields an expression for $\hat{q}$ that involves $\hat{d}$:

$$\hat{q} = \frac{e^{-\hat{d}(p_{S_B}+\hat{p}_N)}(p_{S_B} + \hat{p}_N)[p_{S_W}\hat{p}_R - \hat{p}_{S_W R}(u)]}{p_{S_W}\hat{p}_{S_W R}(u) - p_{S_W}\hat{p}_R(\hat{p}_{S_W} + \hat{f}(u))}. \tag{4.6.2}$$

In (4.6.2), the density estimate $\hat{f}(u)$ is obtained by using the estimates of μ and σ obtained from Algorithm 4.6.1. Evaluation is done at $u = \hat{\mu}$, which is the location of the peak for $p_{S_W R}(u)$. Since we assume a $N(\mu, \sigma)$ delay distribution, we write $\phi_{\hat{\mu},\hat{\sigma}}(\hat{\mu})$ rather than $\hat{f}(\hat{\mu})$ in the remainder of this chapter.

Depending on whether or not the presence of non-linear inhibition is assumed, we develop a corresponding estimation method.

4.6.1 Absence of Non-Linear Inhibition

In the case that $d = 0$, our estimate of p (4.6.1) simplifies to:

$$\hat{p} = 1 - \frac{\hat{p}_R}{(p_{S_B} + \hat{p}_N) + \hat{q}p_{S_W}}. \tag{4.6.3}$$

Furthermore, our expression for $\hat{q}$ (4.6.2) with $d = 0$ reduces to:

$$\hat{q} = \frac{(p_{S_B} + \hat{p}_N)[p_{S_W}\hat{p}_R - \hat{p}_{S_W R}(\hat{\mu})]}{p_{S_W}\hat{p}_{S_W R,h}(\hat{\mu}) - p_{S_W}\hat{p}_R(\hat{p}_{S_W} + \phi_{\hat{\mu},\hat{\sigma}}(\hat{\mu}))}, \tag{4.6.4}$$

where $\hat{p}_{S_W R,h}(\hat{\mu})$ is a nonparametric estimate (4.5.2). We set $\hat{q}$ as zero when evaluation of (4.6.4) yields a negative value.

Because of the dependence of our estimates on nonparametric estimates, the accuracy of our estimates will depend on the choice of bandwidth. As shown in the previous section, the AMISE optimal bandwidth depends on the values of the model parameters. Therefore, to obtain $\hat{p}_{S_W R,h}(\hat{\mu})$ we first find an initial bandwidth h using the already obtained estimates $\hat{\mu}$ and $\hat{\sigma}$, along with $d = 0$, $q = q_0$ and $p = \hat{p}$. Since the error probability is likely relatively small, we use $q_0 = .05$, as an initial value. The estimate of $\hat{p}_{S_W R,h}(\hat{\mu})$ can then be used to obtain $\hat{q}$, which can be used to find $\hat{p}$. If $\hat{q} > 0$, we use our parameter estimates to update the value of h. Using this bandwidth, a nonparametric estimate is obtained once more and then used to obtain a final update of the parameter estimates for q and p.

4.6.2 Presence of Non-Linear Inhibition

In this case $d > 0$, and our expressions for p and q, as given by (4.6.1) and (4.6.2), involve d. Therefore, $\hat{p}$ and $\hat{q}$ can be found upon finding an estimate of d.

The effect of d on the intensity $p_{RR}(u)$ helps us obtain $\hat{d}$. We consider estimation at the trough location $u = 0$. Substituting (4.6.1) and (4.6.2) into the expression for $p_{RR}(0)$ as given in (4.3.4) enables us to express $p_{RR}(0)$ as a univariate function of d. However, a closed-form expression for d cannot be found, so we find an estimate of d by choosing $\hat{d}$ that minimizes the distance between the parametric and nonparametric estimates of $p_{RR}(0)$. This is accomplished by using the `optimize` function in the

statistical package R (R Development Core Team (2006)), which uses a combination of golden section search and successive parabolic interpolation. We use a search interval of $[0, .3]$ for our d estimate.

As in the previous case, since our estimates depend on the nonparametric estimates, we use an initial bandwidth h to obtain parameter estimates. In finding this initial bandwidth, we set d_0 as the midpoint of the search interval for $\hat{d}$. Using the resulting parameter estimates we then update the bandwidth and use this to obtain our final parameter estimates.

4.7 Application to Simulated Data

In this section we study the behaviour of our intensity function estimates for simulated data. Our parameter estimates tend to be close to the true parameter values of the simulated data. Furthermore, plots of the intensity estimates and corresponding nonparametric intensity estimates reveal that the two estimates share similar features.

Recall that since we are assuming a symmetric reaction time distribution, the location of the peaks for the nonparametric intensity estimates of $p_{S_B R}(u)$ and $p_{S_W R}(u)$ should be near the mean reaction times from white and black flashes, respectively. Based on nonparametric intensity estimates of $p_{S_B R}(u)$ and $p_{S_W R}(u)$ for simulated data generated by Algorithm 4.6.1, although the data is generated such that the mean reaction times are μ for both black and white flashes, the two peaks do not always occur at similar locations; the distance between the peak locations was found to vary from 0 to as much as 0.16, with the largest differences occurring when d and/or q are small, and most differences being less than 0.06. The peak location for $p_{S_B R}(u)$ is consistently near μ, while there is more fluctuation in the peak location for $p_{S_W R}(u)$. Examples of peak locations for simulated data with various parameter settings and flash rates are given in Table 4.1.

Table 4.1: Peak locations for nonparametric estimates of $p_{S_W R_W}$ and $p_{S_B R_B}$ for simulated choice RT data, with $\mu = .4$, $\sigma = .08$, $p = .1$, $p_{N_B} = p_{N_W} = 0$, d and q specified in the column headings, and p_{S_W}, p_{S_B} specified by the row names. A bandwidth of $h = .15$ was used. The entries in the table are ($p_{S_B R_B}$ peak location, $p_{S_W R_W}$ peak location).

p_{S_B}, p_{S_W}	$d = 0, q = .05$	$d = .1, q = .05$	$d = .2, q = .05$	$d = .1, q = .02$
.2,.2	(.40,.52)	(.41,.56)	(.41,.45)	(.39,.30)
.2,.2	(.40,.32)	(.41,.42)	(.40,.47)	(.40,.45)
.4,.4	(.40,.31)	(.40,.34)	(.41,.35)	(.40,.33)
.4,.4	(.41,.35)	(.40,.47)	(.39,.46)	(.41,.25)
.6,.6	(.39,.43)	(.40,.27)	(.41,.28)	(.41,.46)
.6,.6	(.40,.49)	(.41,.40)	(.40,.47)	(.41,.31)
.5,.7	(.39,.49)	(.39,.35)	(.42,.36)	(.42,.34)
.5,.7	(.39,.36)	(.40,.50)	(.41,.43)	(.38,.43)

4.7.1 Absence of Non-Linear Inhibition

Here, we provide two examples of estimation under the assumption of $d = 0$, for simulated data with $d = 0$. Estimates are obtained using the method of Section 4.6.1 with Algorithm 4.6.1.

In our first example, we consider simulated data with equal flash rates. The simulated data consists of 800 flashes and has parameters $\mu = .4$, $\sigma = .1$, $p_N = 0$, $d = 0, p = .15, p_{S_B} = .5, p_{S_W} = .5$, and $q = .1$. The estimates we obtain are $\hat{\mu} = 0.394$, $\hat{\sigma} = 0.093$, $\hat{q} = 0.098$, $\hat{p} = 0.163$, and $\hat{p}_N = 0.019$. An illustration comparing the parametric and nonparametric intensity estimates is given in Figure 4.3. The general features of the nonparametric estimates appear to be captured by our model.

The means and standard errors of estimates based on 500 simulations of 800 flashes with the same parameter settings as in the previous example are given in Table 4.2 The estimates have small standard errors with the largest standard error, .0434 belonging to p.

Next, we consider simulated data with different flash rates. The simulated data consists of 800 flashes and has parameters $\mu = .4$, $\sigma = .1$, $d = 0$, $p = .15$, $p_{S_B} = .7$, $p_{S_W} = .5$, $q = .1$, and $p_N = 0$. The estimates we obtain are $\hat{\mu} = 0.395$, $\hat{\sigma} = 0.092$, $\hat{q} = 0.097$, $\hat{p} = 0.188$, and $\hat{p}_N = 0.037$. An illustration comparing the parametric

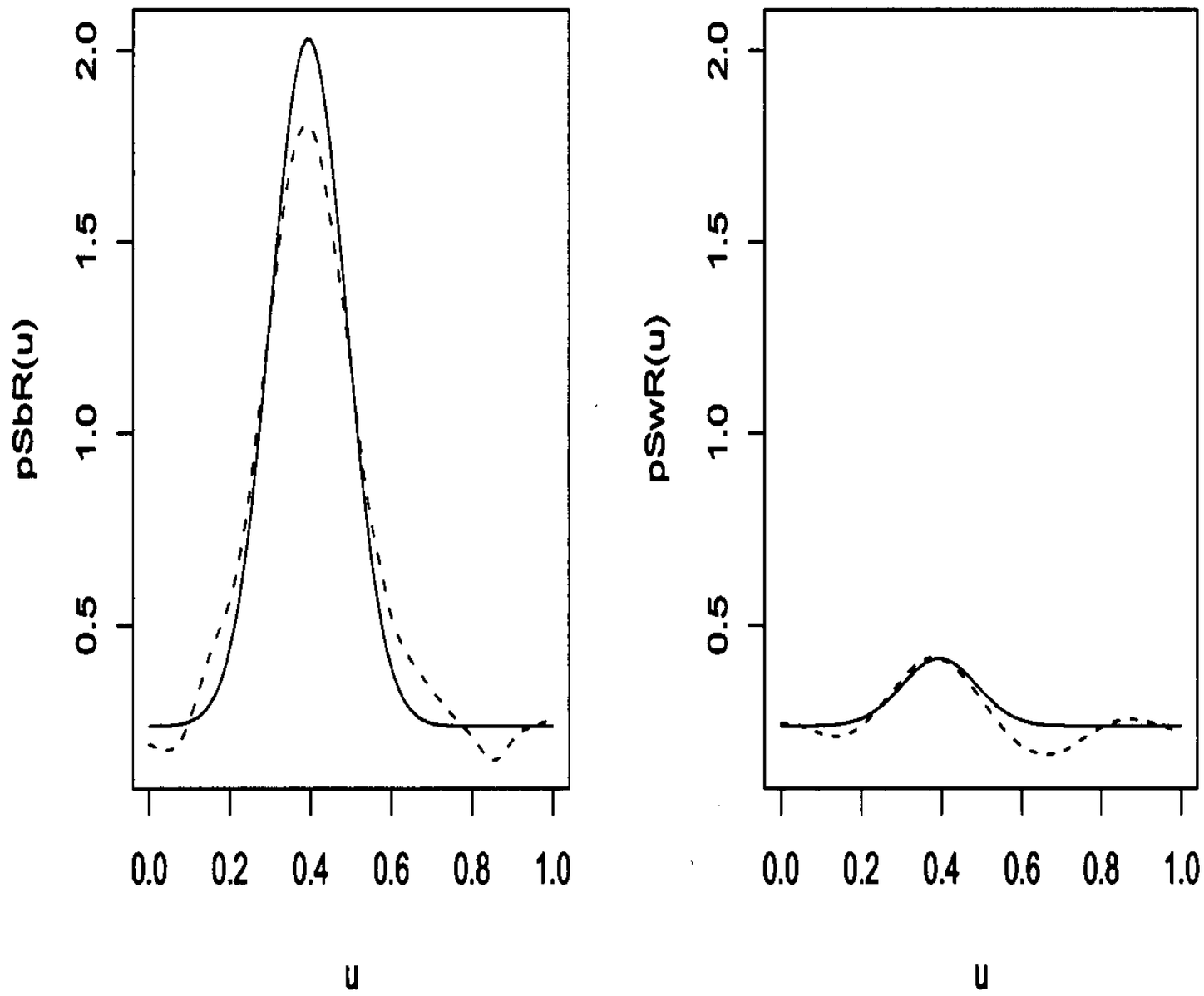


Figure 4.3: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .5$, and $d = 0$.

and nonparametric intensity estimates is given in Figure 4.4. The nonparametric and parametric estimates exhibit the same main qualitative features; the nonparametric estimates are affected by greater variability

Table 4.3 displays the standard errors of the estimates for 500 simulations with parameters $\mu = .4, \sigma = .1, d = 0, p = .15, p_{S_B} = .7, p_{S_W} = .5, p_N = 0$, and $q = .1$. The estimates of p an q have the largest standard errors, .0380 and .0306, respectively.

4.7.2 Presence of Non-Linear Inhibition

Here, we provide two examples of estimation for simulated data with $d = .2$. Estimates are obtained using the method of Section 4.6.2 with Algorithm 4.6.1.

Table 4.2: Standard errors of parameter estimates for simulated data with rate $p_{S_B} = p_{S_W} = .5$ and $d = 0$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.398	.0066
σ	.1	.091	.0038
q	.1	.074	.0268
p	.15	.159	.0434
p_N	0	.022	.0052

Table 4.3: Standard errors of parameter estimates for simulated data with $p_{S_W} = .7$, $p_{S_B} = .5$, and $d = 0$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.398	.0060
σ	.1	.089	.0035
q	.1	.076	.0306
p	.15	.175	.0380
p_N	0	.040	.0078

In our first example, we consider simulated data with equal flash rates. The simulated data consists of 800 flashes and has parameters $\mu = .4$, $\sigma = .1$, $d = .2$, $p = .15$, $p_{S_B} = .5$, $p_{S_W} = .5$, and $q = .1$. The estimates we obtain are $\hat{\mu} = 0.398$, $\hat{\sigma} = 0.087$, $\hat{d} = .193$, $\hat{q} = 0.083$, $\hat{p} = 0.183$, and $\hat{p}_N = 0.011$. An illustration comparing the parametric and nonparametric intensity estimates is given in Figure 4.5. The same main qualitative features are exhibited in both estimates.

Table 4.4 displays the standard errors of the estimates for 500 simulations with parameters $\mu = .4, \sigma = .1, d = .2, p = .15, p_{S_B} = .5, p_{S_W} = .5, p_N = 0$, and $q = .1$. The estimates of d have the largest standard error, .0468, and tend to be underestimated.

Next, we consider simulated data with different flash rates. The simulated data consists of 800 flashes and has parameters $\mu = .4$, $\sigma = .1$, $d = .2$, $p = .15$, $p_{S_B} = .7$, $p_{S_W} = .5$, $q = .1$, $p_N = 0$. The estimates we obtain are $\hat{\mu} = 0.406$, $\hat{\sigma} = 0.090$, $\hat{d} = .214$, $\hat{q} = 0.100$, $\hat{p} = 0.155$, and $\hat{p}_N = 0.007$. The parametric and nonparametric intensity estimates are displayed in Figure 4.6. The nonparametric estimates have

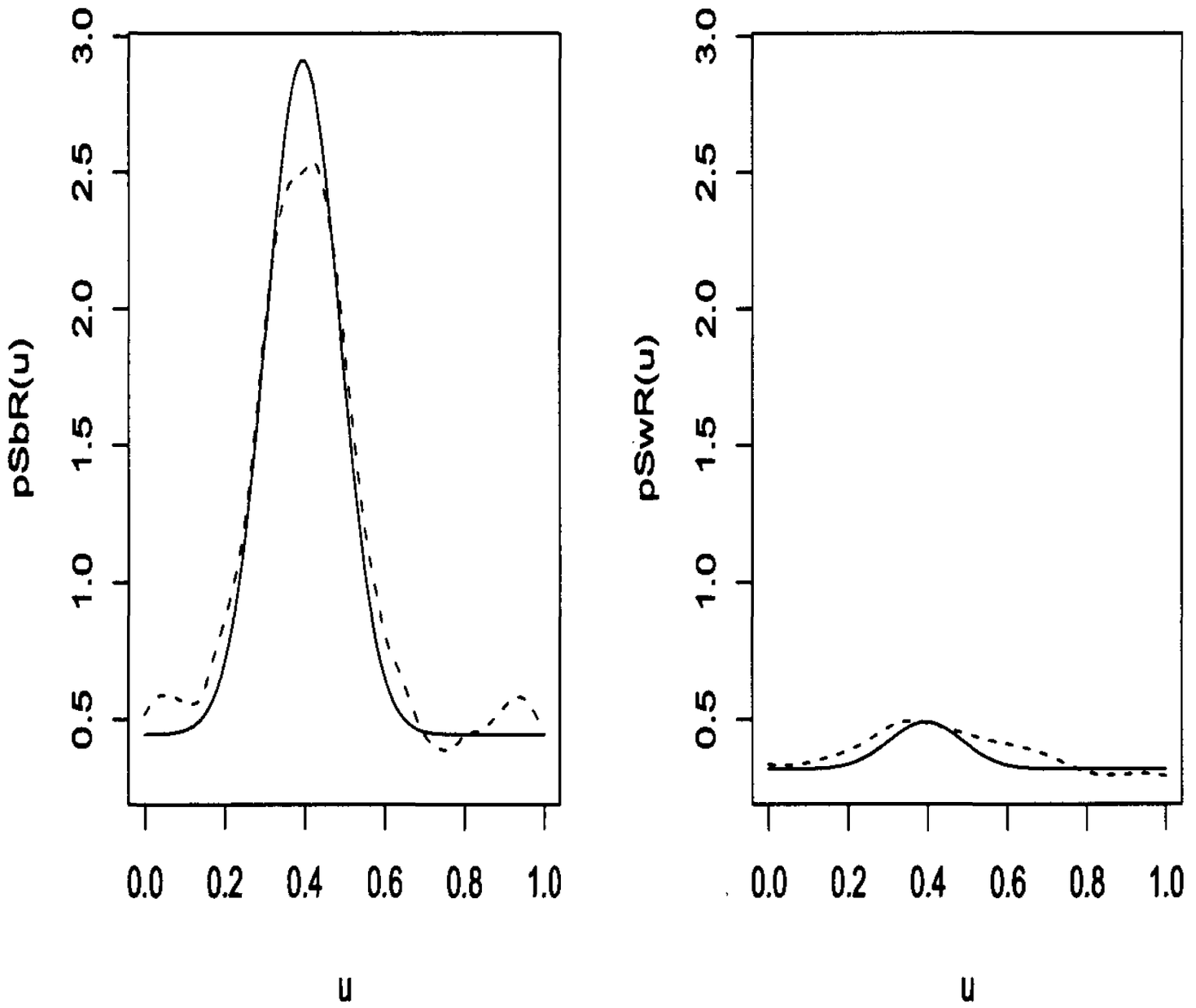


Figure 4.4: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .7$, and $d = 0$.

greater variability, but both estimates have similar characteristics.

Standard errors of the estimates for 500 simulations with parameters $\mu = .4, \sigma = .1, d = .2, p = .15, p_{S_B} = .7, p_{S_W} = .5, p_N = 0$, and $q = .1$ are given in Table 4.5. The standard errors are quite small with $\hat{p}$ having the largest standard error .0491.

4.8 Model Fitting for Go-no go RT data

We fit our model to ten pooled runs (pooling based on 8 runs of 100 flashes each), each with equal stimulus rates. A description of how we pooled the data is provided in Section 3.3.6. We also consider an example of estimation for the individual unpooled data. The stimulus rates are 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 2.0, 4.0, 5.0 and 8.0; because

Table 4.4: Standard errors of parameter estimates for simulated data with rate $p_{S_B} = p_{S_W} = .5$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.402	.0066
σ	.1	.092	.0086
d	.2	.189	.0513
q	.1	.074	.0288
p	.15	.139	.0511
p_N	0	.009	.0033

Table 4.5: Standard errors of parameter estimates for simulated data with $p_{S_W} = .7$, $p_{S_B} = .5$, and $d = 0$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.404	.0065
σ	.1	.090	.0037
d	.2	.196	.0388
q	.1	.069	.0333
p	.15	.138	.0491
p_N	0	.013	.0043

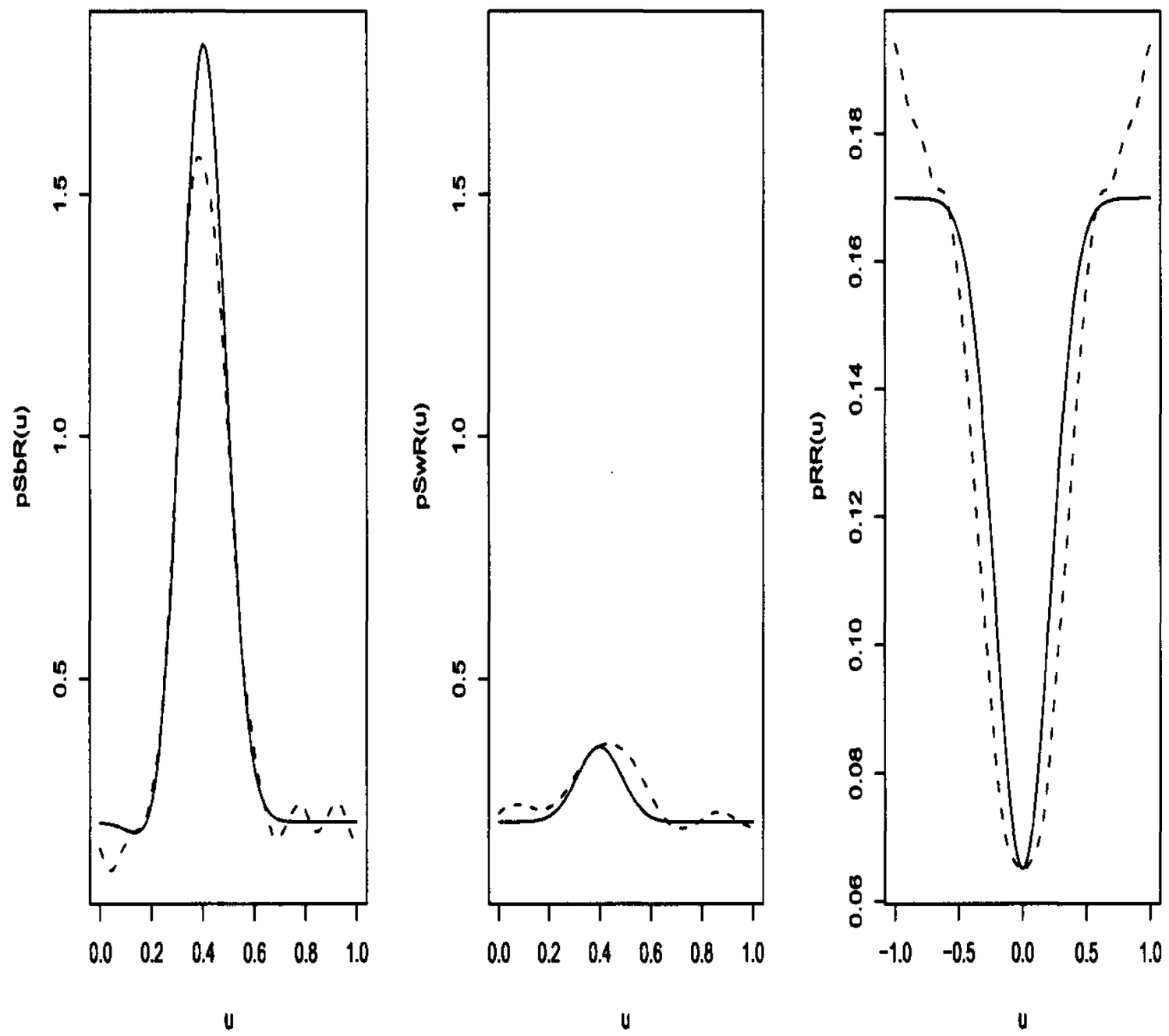


Figure 4.5: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .5$, and $d = .2$.

of equal stimulus rates a rate of 1.0 corresponds to $p_{S_B} = p_{S_W} = 0.5$.

We make the simplifying assumption that the distribution of the reaction times from white and from black flashes are identical, so that they have the same parameters μ, σ (Assumption 3). We use Algorithm 4.6.1 to obtain the estimates of μ, σ, and p_N. For the real data we do not make the restriction of assuming $d = 0$. Thus, estimates of p, q, and d are found using the method of Section 4.6.2.

Recall that the location of the peaks for the nonparametric intensity estimates of $p_{S_B R}(u)$ and $p_{S_W R}(u)$ should be near the mean reaction times from white and black flashes, respectively. Based on plots of these nonparametric estimates for the ten pooled data sets the assumption of equal means for the black and white reaction

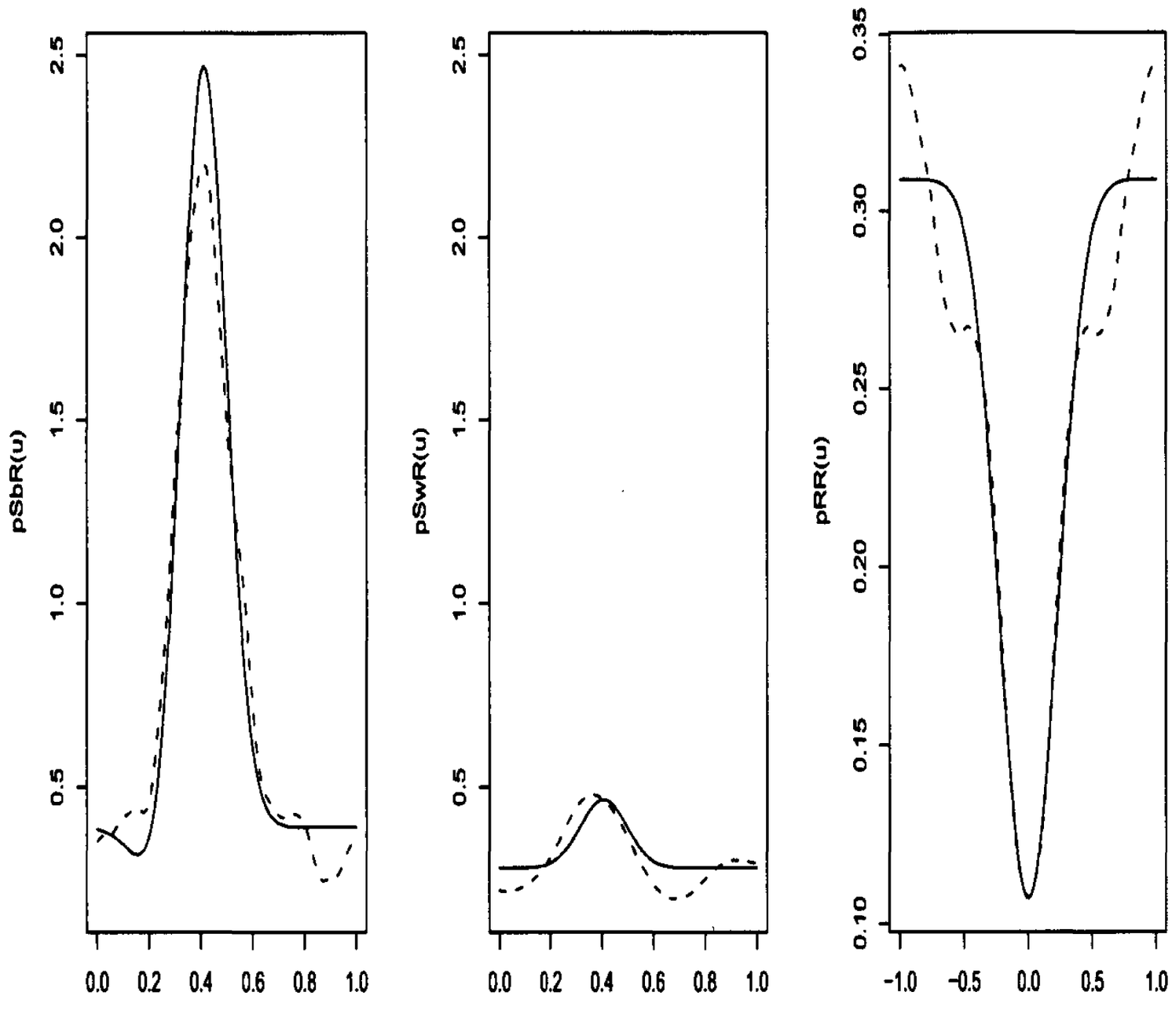


Figure 4.6: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .7$, and $d = .2$.

times appears to be somewhat unrealistic for the majority of the data sets. However, we also found that for simulated data there is large variability in the location of the peak for $p_{S_W R}(u)$. Taking this observation into account, the assumption of equal RT means from black and white flashes is not unreasonable. Locations of the peaks for the ten pooled data sets are given in Table 4.6, while the parameter estimates are given in Table 4.7.

The mean reaction time estimates $\hat{\mu}$ are in the interval (0.3,0.4), and tend to decrease as the rates increase. A faster reaction time as the rate increases may result from the high flash intensity which makes the subject expect flashes with a high frequency. In simple RT experiments run at nine of the ten rates the mean reaction

Table 4.6: Peak locations for nonparametric estimates of $p_{S_B R}$ and $p_{S_W R}$ for pooled choice RT data, using a bandwidth of $h = .15$.

p_{S_B}, p_{S_W}	$p_{S_B R}(u)$ peak location	$p_{S_W R}(u)$ peak location
0.2	.39	.31
0.3	.37	.31
0.4	.36	.32
0.5	.36	.31
0.6	.36	.26
0.7	.33	.29
1.0	.33	.29
2.0	.33	.26
2.5	.30	.27
4.0	.30	.24

Table 4.7: Parameter estimates for each of the ten pooled data sets

p_{S_B}, p_{S_W}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{d}$	$\hat{p}$	$\hat{q}$	$\hat{p_N}$
0.2	.393	.077	7×10^{-5}	.114	.061	.015
0.3	.376	.068	7×10^{-5}	0	.024	.032
0.4	.376	.082	.159	0	.023	.014
0.5	.368	.073	7×10^{-5}	.054	.012	.032
0.6	.363	.074	.123	0	0	.027
0.7	.349	.088	.199	0	.041	.023
1.0	.339	.069	.098	0	.040	.061
2.0	.327	.062	.170	0	.044	.074
2.5	.308	.057	.175	0	.014	.069
4.0	.307	.051	.107	.112	.0004	.131

times are near 0.3 (see Table 3.2 for the complete results), so the mean is higher for go-no go RT experiments. That is, the time to respond to a flash in a go-no go RT experiment is longer than in a simple RT experiment, which seems reasonable since a go-no go RT experiment involves the extra tasks of identifying and responding to only the black flashes. In psychophysics, this result is well-known (see Luce (1986)).

The rate of the noise process also increases with the stimulus rates, which seems to agree with intuition because as flashes are presented with a higher frequency it seems more likely that the subject will be overwhelmed and press the button in the absence of a flash.

The estimates of d are essentially zero when both flash rates are .2, .3, or .5, while $\hat{d} > .09$ for the remaining stimulus rates. This suggests that temporal summation (two consecutive flashes viewed as one bright flash) occurs more frequently in the visual system when flashes are presented at a fast rate. Recall that in our parametric model, when a black flash occurs within d time units after the previous black flash, the two flashes are perceived as one bright flash, and only one response occurs.

Completely random thinning does not appear to occur, as indicated by $\hat{p} = 0$ for the majority of the cases. For most data sets, the estimated error probability $\hat{q}$ is small and positive, giving an indication of the proportion of wrong response types.

Comparisons of the parametric and nonparametric second-order intensity estimates reveal that the correspondence between the two curves is closest for low to moderate flash rates (e.g. rates of .2 to .7 flashes/s for each flash type). For the three data sets with $\hat{d} = 0$ ($p_{S_B}, p_{S_W} \in \{0.2, 0.3, 0.5\}$) we do not consider plots of $p_{RR}(u)$ because the intensity is constant. In Figures 4.7 to 4.16 it can be seen that key features of the nonparametric estimates are captured by the parametric estimates.

In Figures 4.7 and 4.8 the difference in the peak locations is quite apparent, as observed in Table 4.7. For these data sets we also have $\hat{d}$ near zero. Such results are consistent with the results obtained for our simulations; for simulated data with small d values there were occurrences in which the difference in peak locations was as large as 0.16.

Comparison plots of the intensity estimates for go-no go RT data with black and

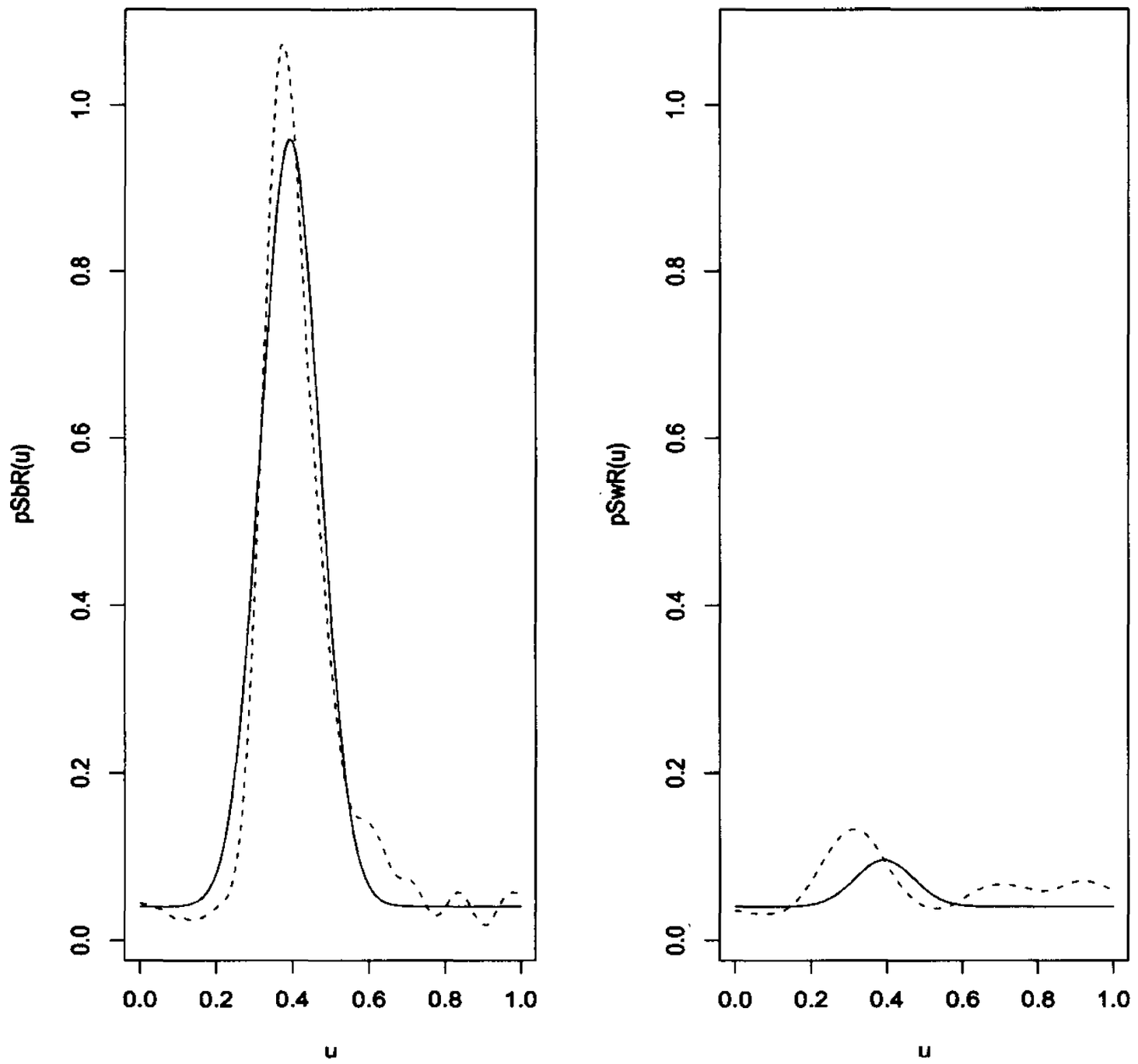


Figure 4.7: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = .2$.

white flash rates of $p_{S_B} = p_{S_W} = .7$ are given in Figure 4.12. In the plot for $p_{S_B R}(u)$ there is a dip to the left of the peak in both estimates, and the peak location is similar for the two estimates. The nonparametric estimate of $p_{S_W R}(u)$ has a lower peak location than the parametric estimate, but the model is able to capture the low height of the intensity. Both estimates of $p_{RR}(u)$ reach the same trough depth, with the parametric estimate slightly wider.

As a comparison, we also studied parameter estimation for the unpooled data. The parameter estimates for the unpooled data tend to fluctuate around the estimates obtained for the pooled data. As an example, the parameter estimates for the unpooled data sets with $p_{S_B} = p_{S_W} = .7$ have the means and variances as given in

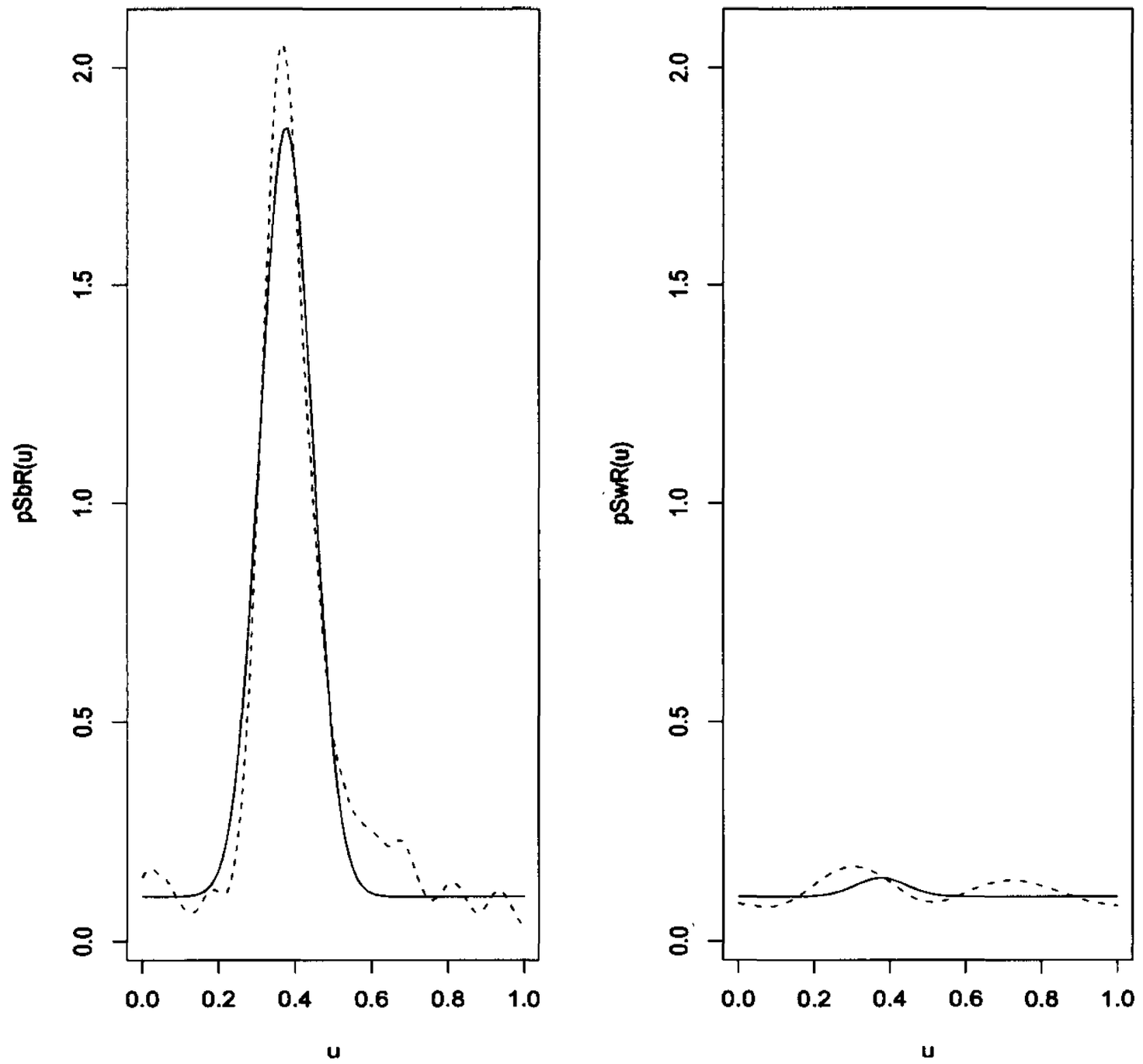


Figure 4.8: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = .3$.

Table 4.8, where the pooled estimates are provided as well. The means of the parameter estimates for the unpooled data are quite close to the estimates obtained for the pooled data.

4.9 A Go-no go RT Threshold Model

In a simple RT experiment, if two flash types are presented, the subject presses a button immediately after perceiving either flash type. In Chapter 3 we assumed that there was no bias in response to flash type, so that flash type was not taken into account in the corresponding threshold model. However, since only black flashes are responded to in a go-no go RT experiment, flash type must be taken into consid-

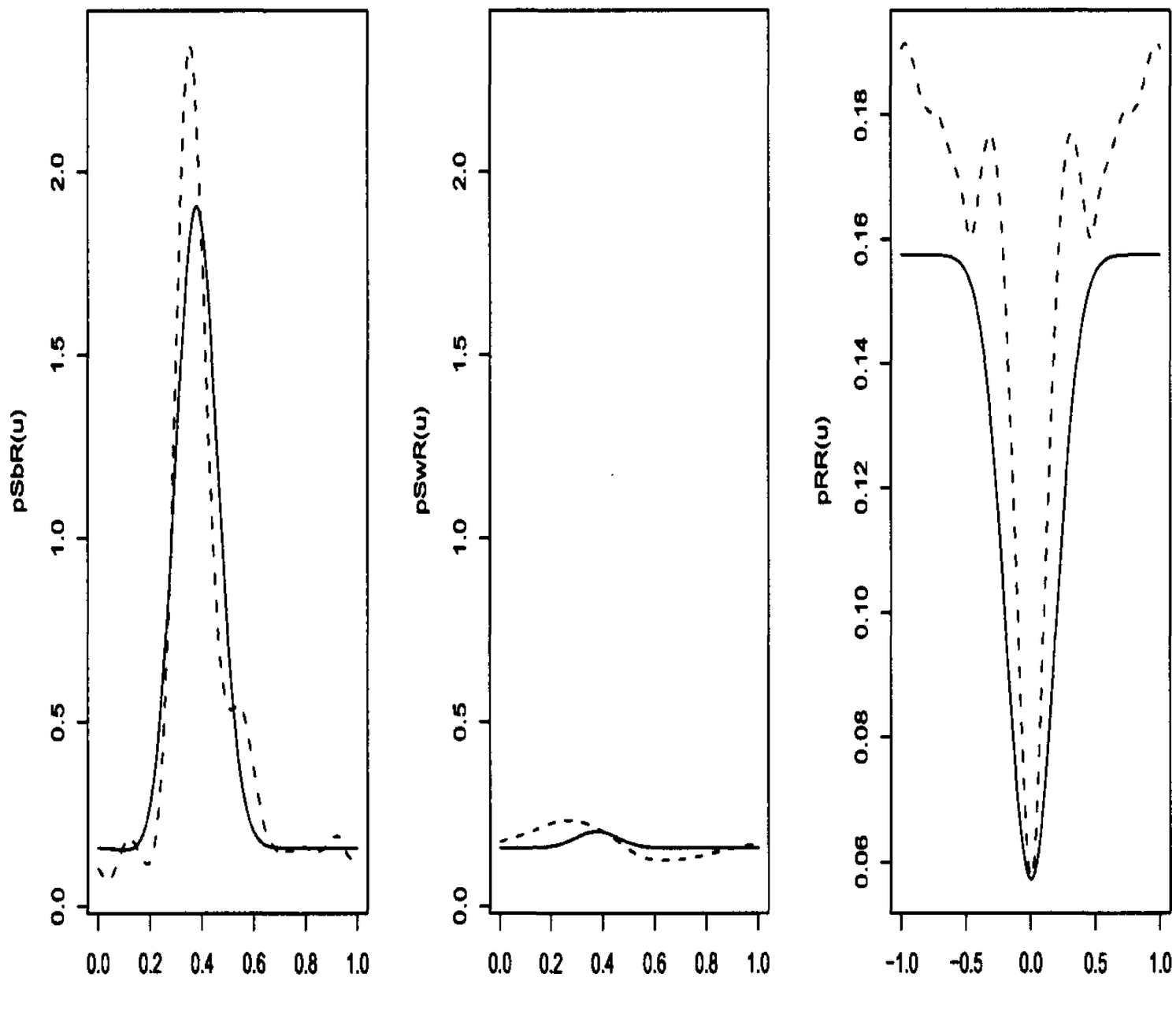


Figure 4.9: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = .4$.

eration; most responses occur after a black flash, and those in response to a white flash are errors. As in the simple RT threshold model the go-no go threshold model involves one internal potential and a single threshold.

In our conceptual model, the stimulus, which consist of both black and white flashes, passes through an internal filter. Internal noise is added to the filter output, forming the internal potential. When the internal potential exceeds the threshold θ the subject decides that there is a black flash and presses the button.

Prior to setting up the threshold model, notation must be introduced, some of it repeated from Chapter 3 for convenience. We discretize a point process S into the 0-1 time series X as follows, where we choose some lag value m to create the time intervals

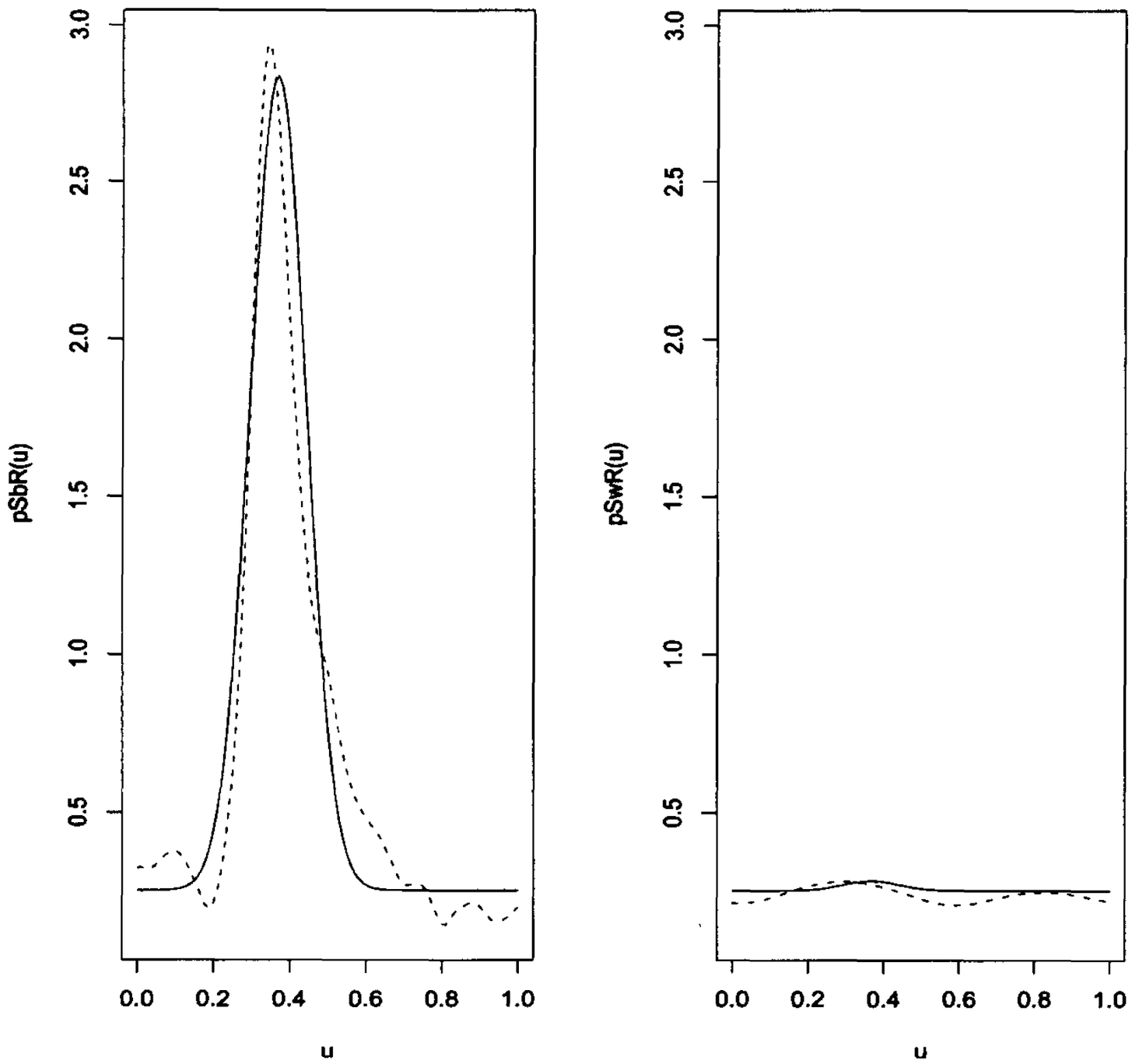


Figure 4.10: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = .5$.

$\{(0, m], (m, 2m], \ldots, (T, T+m]\} = \{I_1, I_2, \ldots, I_T\}$, where $T = m\lceil \max(R^B, R^W)/m \rceil$

$$X_j = \begin{cases} 1, & \text{if there exists } S_k \text{ such that } S_k \in I_j \\ 0, & \text{otherwise} \end{cases} ;$$

We use the following notation:

- S^B and S^W denote the black and white stimulus point processes, respectively;
- R denotes the response point process;
- X^B and X^W denote the discretized black and white stimulus point processes;
- Y denotes the discretized response process;

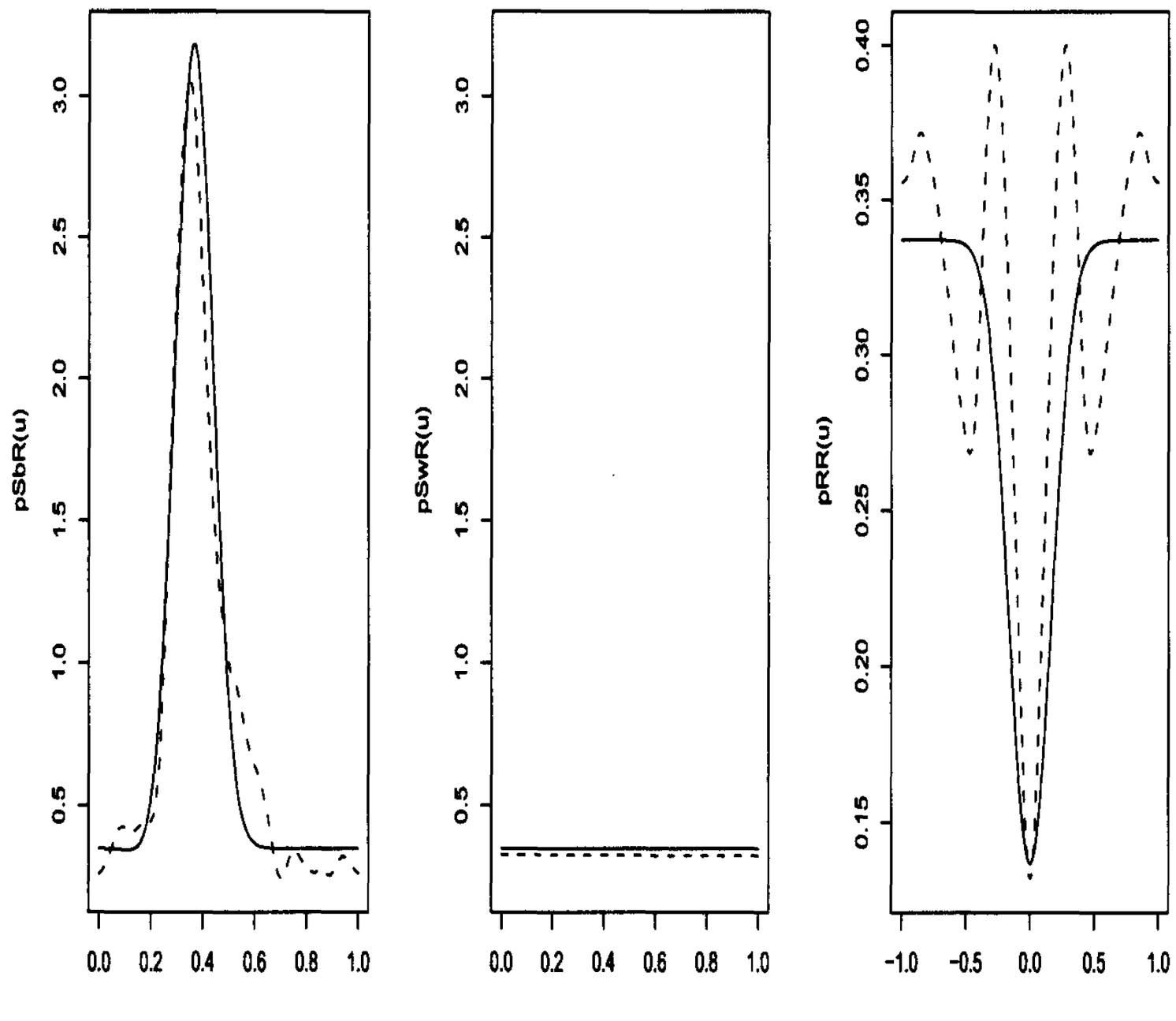


Figure 4.11: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = .6$.

- γ_j denotes as of time t, the number of time intervals elapsed since the last interval containing a response ($t \in I_j$, so that j is the index of the interval containing time t);

- $a_{t-\tau}$ represents the effect of a black flash on the internal potential at time t, assuming such a flash occurs in a time interval with right endpoint τ;

- $b_{t-\tau}$ represents the effect of a white flash on the internal potential at time t, assuming such a flash occurs in a time interval with right endpoint τ.

- U_t is the internal potential at time t.

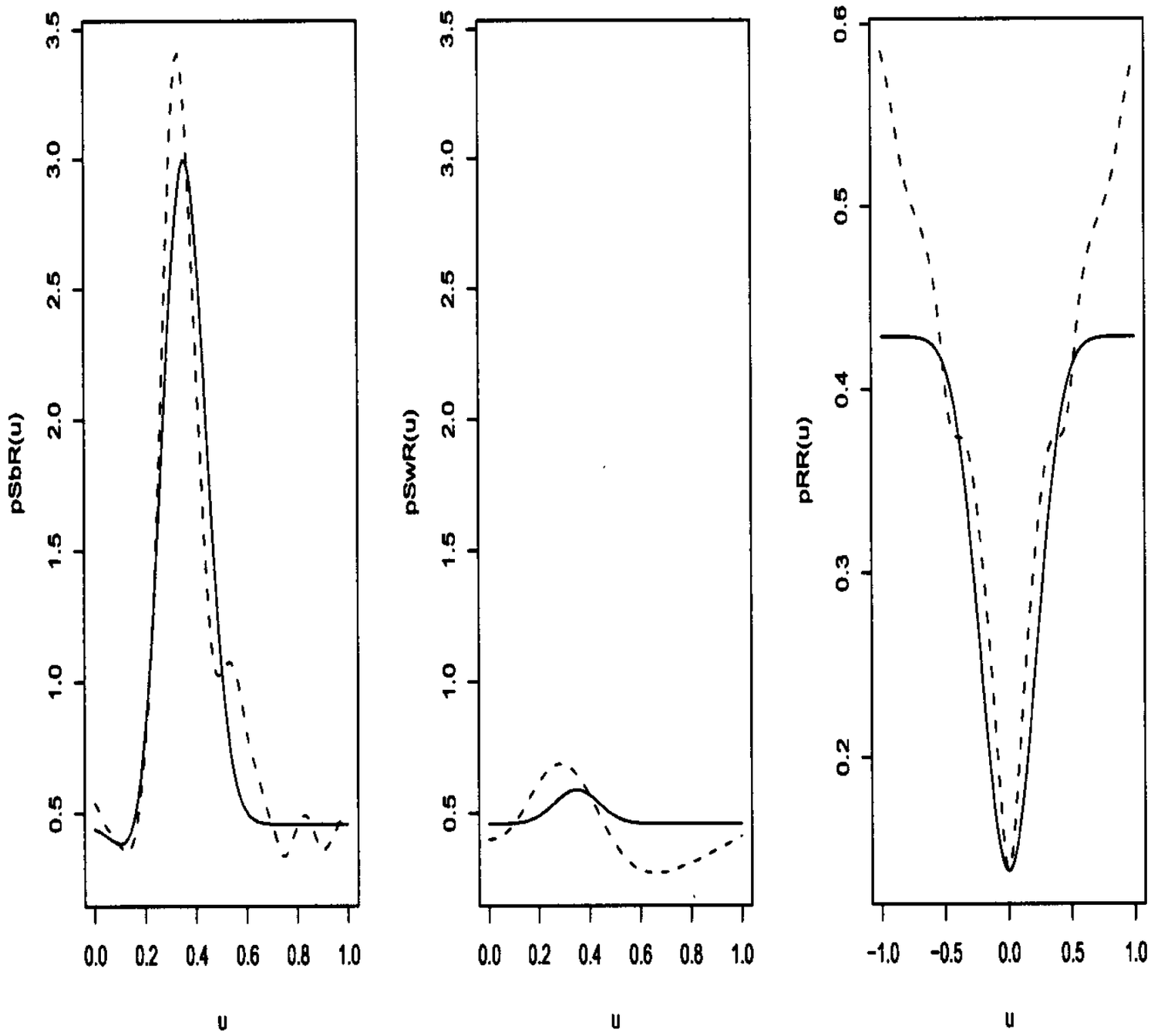


Figure 4.12: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = .7$.

A response occurs when

$$U_t + \varepsilon_t > \theta,$$

where θ is the threshold, and ε_t is internal noise, which we assume to have a symmetric distribution about 0.

A simple model for the internal potential does not include interaction terms and is given by

$$U_t = \sum_{u=0}^{g_t - 1} (a_u X^B_{t-u} + b_u X^W_{t-u}),$$

where $g_t = \min\{\gamma_t, G\}$, and G is the maximum number of time intervals (prior to the current time interval) that are of interest. Again, we consider only lags of less than 1 s; when $m = .05$ we would have $G = 20$.

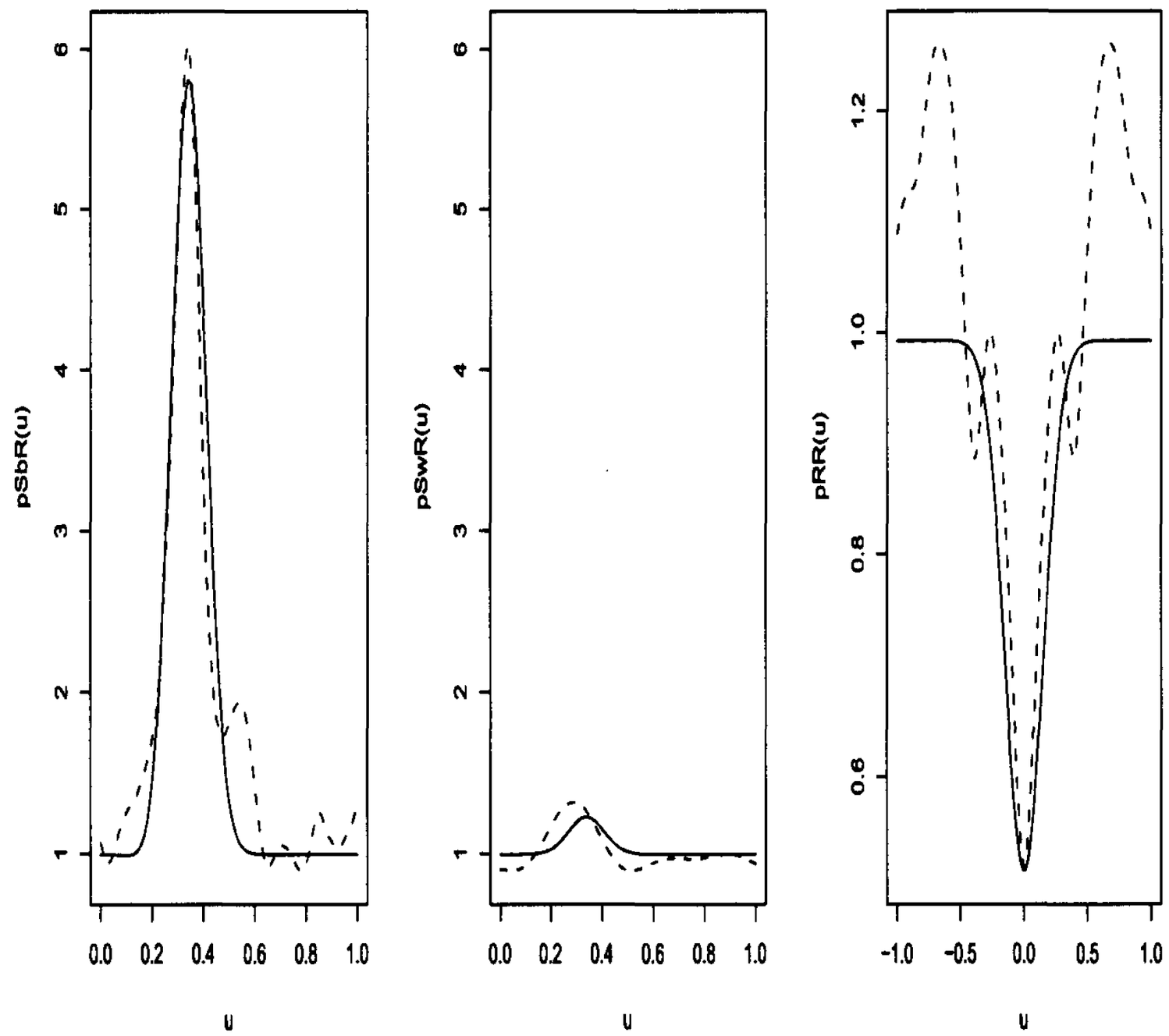


Figure 4.13: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = 1.0$.

We consider interactions of a black flash occurrence in a time interval with right endpoint t with the occurrence of any black flashes occurring within d time units earlier, for some d. For simplicity we set $d = rm$, for some integer r, since we are working with data that is discretized using lag m. Letting $Z_j = I\{\sum_{i=1}^{r} X_{j-i}^B > 0\}$, the internal potential can be expressed as

$$U_t = \sum_{u=0}^{g_t-1}(a_u X_{t-u}^B + b_u X_{t-u}^W) + \sum_{u=0}^{z_t-1} c_u X_{t-u}^B Z_{t-u}^B,$$

where $z_t = \min\{\gamma_t, z\}$, and $z \leq G$ is the maximum number of time intervals (prior to the current time interval) that are of interest for an interaction effect on the internal potential.

The coefficient of the interaction terms c_u can be used to make inferences on any

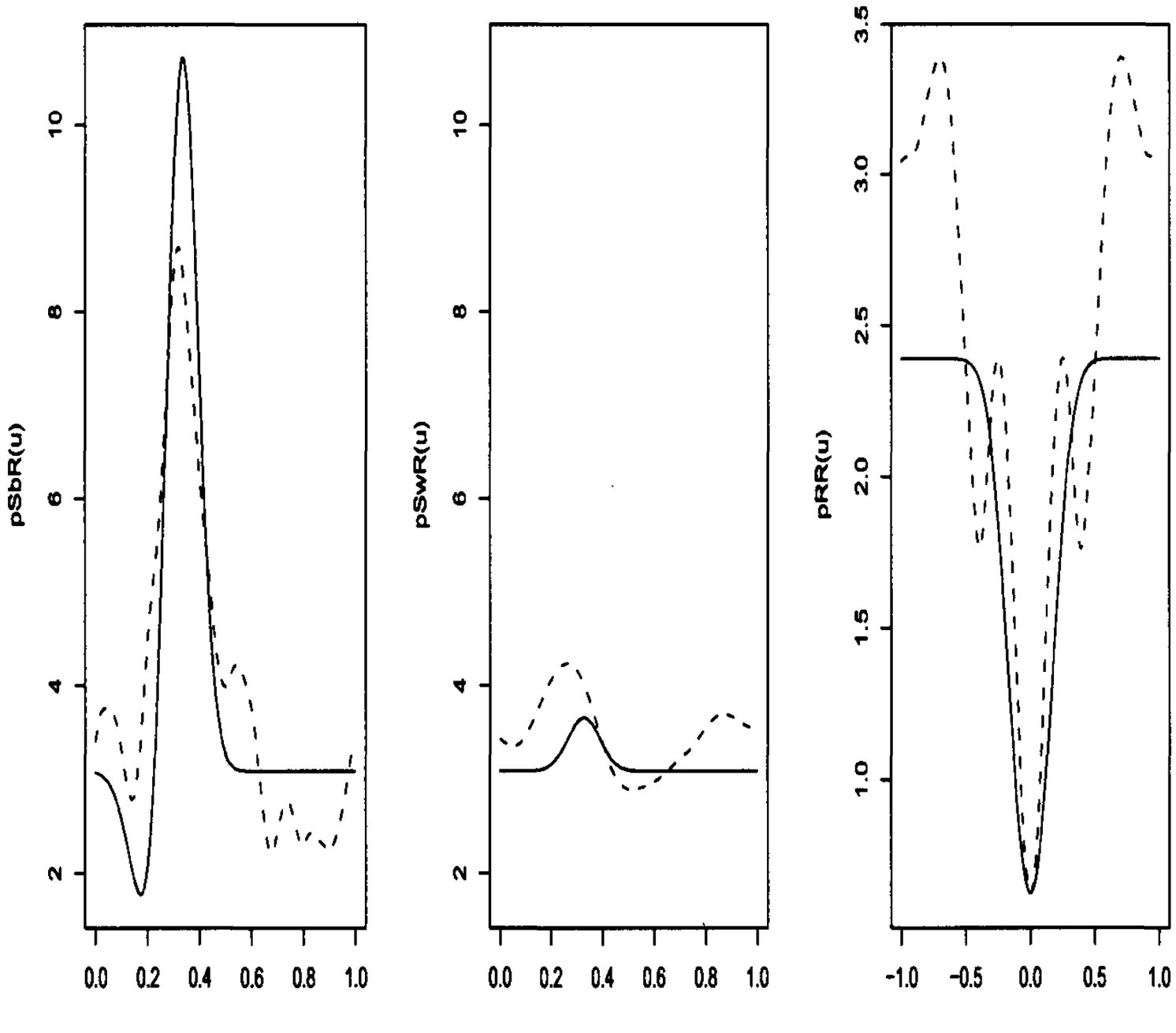


Figure 4.14: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = 2.0$.

inhibition or facilitation occurring among the black flashes. If c_u is significantly different from zero and is positive then the term $c_u X^B_{t-u} Z^B_{t-u}$ gives a positive contribution to the internal potential U_t when there is a black flash at time $t-u$ and another black flash within d time units earlier. With an increase in U_t, the internal potential is closer to crossing the threshold θ and causing a response. Thus, positive significant interaction coefficients provide evidence of facilitation among the black flashes. By similar reasoning it can be seen that negative coefficients are indicative of inhibition.

Our fitted model has the form

$$\hat{Y}_t = \begin{cases} 0, & \text{if } \hat{U}_t < \hat{\theta} \\ 1, & \text{if } \hat{U}_t > \hat{\theta} \end{cases},$$

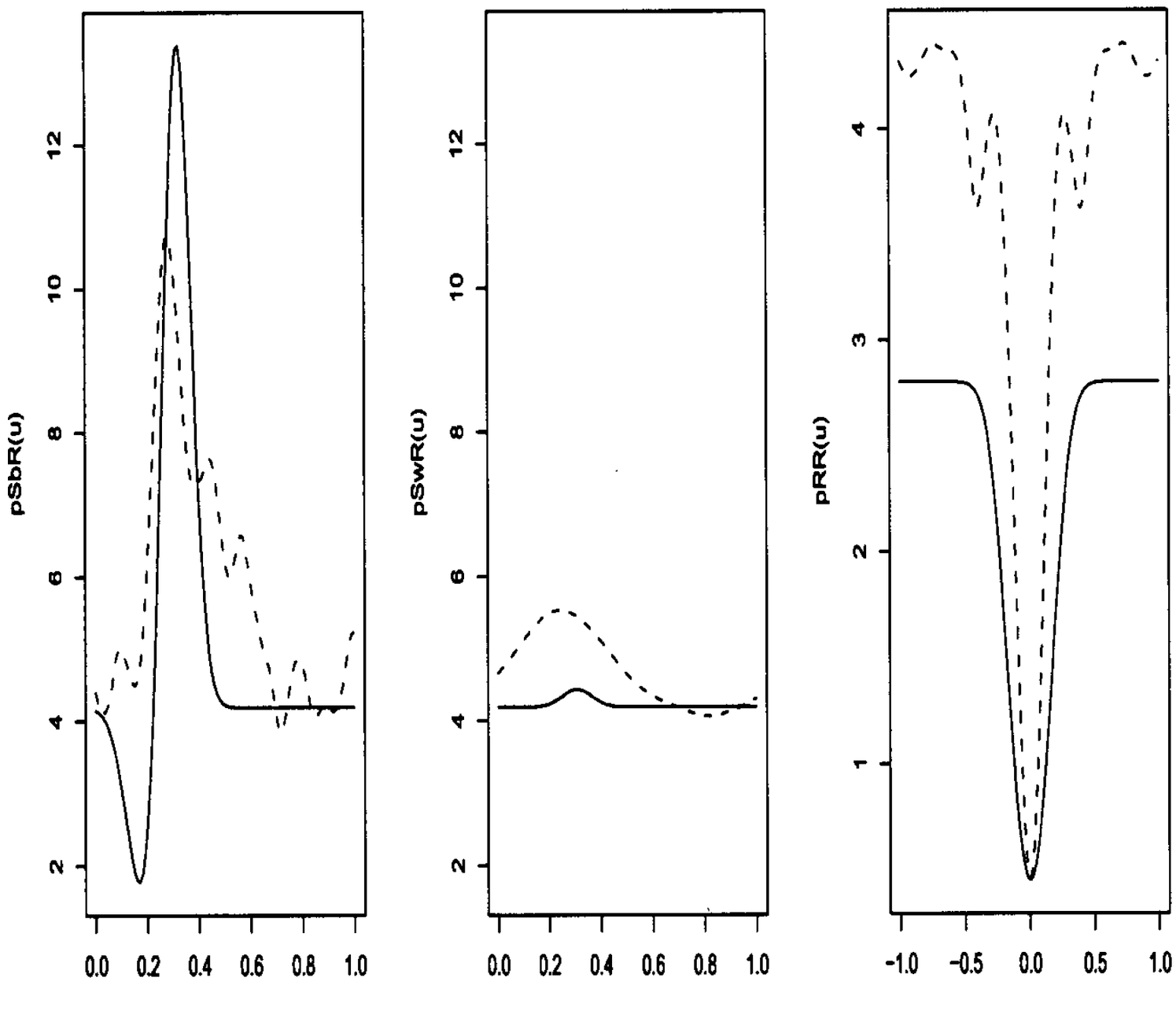


Figure 4.15: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = 2.5$.

where $\hat{U}_t$ is a function of the stimuli and the coefficient estimates.

We fit the model

$$\hat{Y}_t = \begin{cases} 0, & \text{if } \hat{Y}_t^* < 0 \\ 1, & \text{if } \hat{Y}_t^* > 0 \end{cases},$$

where

$$\hat{Y}_t^* = \hat{U}_t + \hat{\beta}$$

so that the estimate of the threshold $\hat{\theta}$ is the negative of the intercept estimate β.

We fit models to ten go-no go pooled RT data sets (pooling based on 8 runs of 100 flashes each), each with equal stimulus rates. We pool the data as described in Section 3.3.6. The rates of flashes per second are 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 2.0, 4.0,

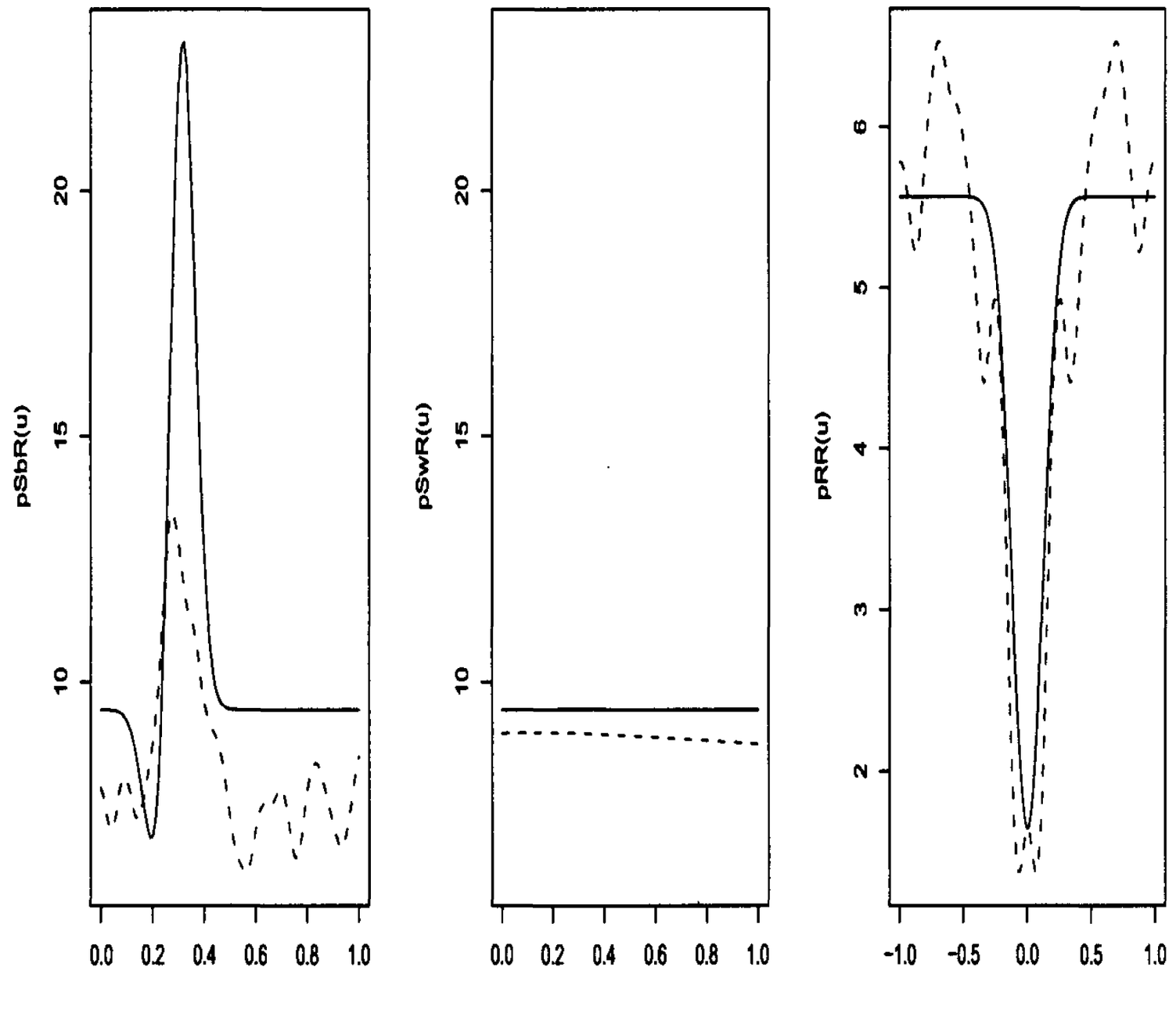


Figure 4.16: Second-order parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled go-no go RT data with $p_{S_B} = p_{S_W} = 4.0$.

5.0 and 8.0. In each experiment black and white flashes are presented with equal rates so that each has a rate equal to half of the total flash rate. When converting the data to 0-1 time series a lag of $m = .05$ is used. We employ the same model assessment strategy as we did in Chapter 3.

In order to study the behaviour of the fitted model and diagnostic for data in which we know the relationship between the flashes and responses, we fit our model to data generated from our parametric go-no go model using Algorithm 4.2.1.

Based on the diagnostic plots for simulations resulting from various parameter settings, it appears that regardless of the presence or absence of thinning within the black flashes (parametric d parameter non-zero or zero, respectively) the fit tends to

Table 4.8: Comparison of pooled and unpooled parameter estimates for data with rate $p_S = .7$

parameter	pooled estimate	mean of unpooled estimates	standard deviation of unpooled estimates
μ	.349	.354	.0008
σ	.088	.083	.0002
d	.199	.144	.0066
q	.041	.049	.0046
p	0	.017	.0022
p_N	.023	.021	.0001

improve as the flash rate increases. Such an improvement is usually most apparent when the two flash rates are at least 0.5. Any problems in the fit occur in the upper tail, $U_t > \hat{\theta}$, where empirical probabilities tend to be much smaller than those resulting from the model.

An example of a close fit is for a simulation with parameters $\mu = .5$, $\sigma = .12$, $d = .1$, $p = .15$, $p_{S_B} = 2$, $p_{S_W} = 2$ and $q = .1$. This can be seen in the diagnostic plot, which is provided in Figure 4.17. Figure 4.18 displays plots of the correct (black flash-black response) and error (white-flash-black response) linear filters. As expected, a peak for the correct impulse is located near 0.5, the same value as the mean parameter μ. Before 0.2 s there appears to be an increase in the number of errors resulting from a response to a white flash, as can be seen from the error impulse peak before $u = .2$.

The majority of the interaction coefficients are negative, agreeing with the form of the parametric model, which incorporates a thinning mechanism.

We now discuss the fitted threshold models for the experimental data. The "best-fitting" model varies for each of the data sets, with different d parameters, and different numbers of interaction terms. The threshold estimates are given in Table 4.9, from which it can be seen that the threshold tends to decrease with the stimulus rate. Such behavior is expected since the response frequency should increase as the flash rate increases. The internal potential U_t will cross the threshold more frequently when there is a fast stimulus rate; in a short time interval more flashes occur so that after a small amount of time a response occurs. In order for the internal potential to

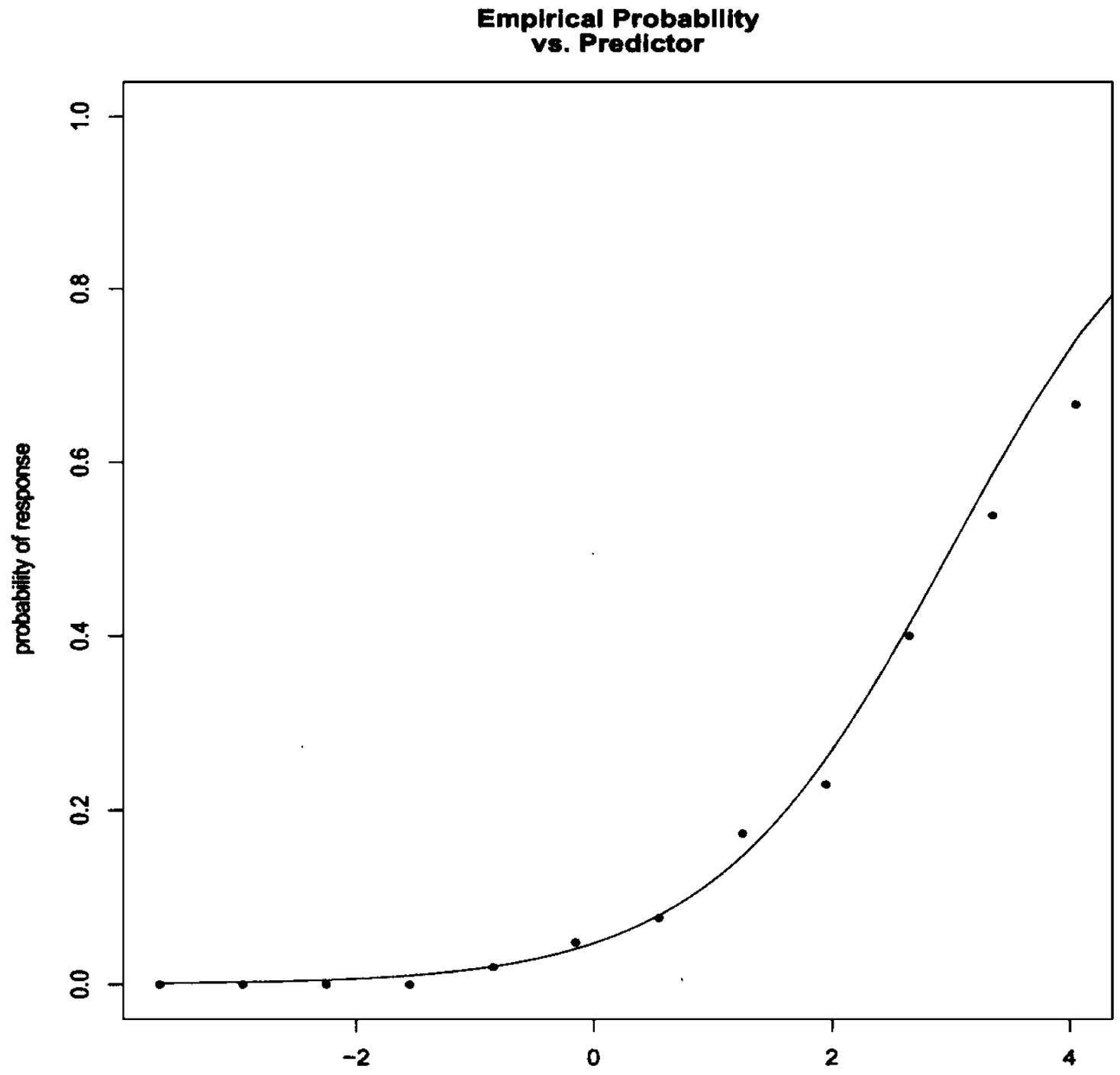


Figure 4.17: Diagnostic plot for simulated go-no go RT data with parameters $\mu = .5$, $\sigma = .12$, $d = .1$, $p = .15$, $p_{S_B} = 2$, $p_{S_W} = 2$ and $q = .1$. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

cross the thresholds within a short period of time it is anticipated that the threshold must not be large when the flash rate is high.

Plots of the correct (response to black stimuli) and error (response to white stimuli) impulse responses are provided in Figures 4.19 and 4.20, while plots of the peaks of each impulse as function of the lag are displayed in Figures 4.21 and 4.22. The peak locations are between 0.3 and 0.4, which agree with our estimates of the mean reaction time using our parametric model.

There is a tendency for the peaks of the correct response impulses to decrease with the stimulus rate, which is consistent with intuition. With a slower stimulus rate one is less likely to miss a stimulus or mistakenly press the wrong button. A higher peak

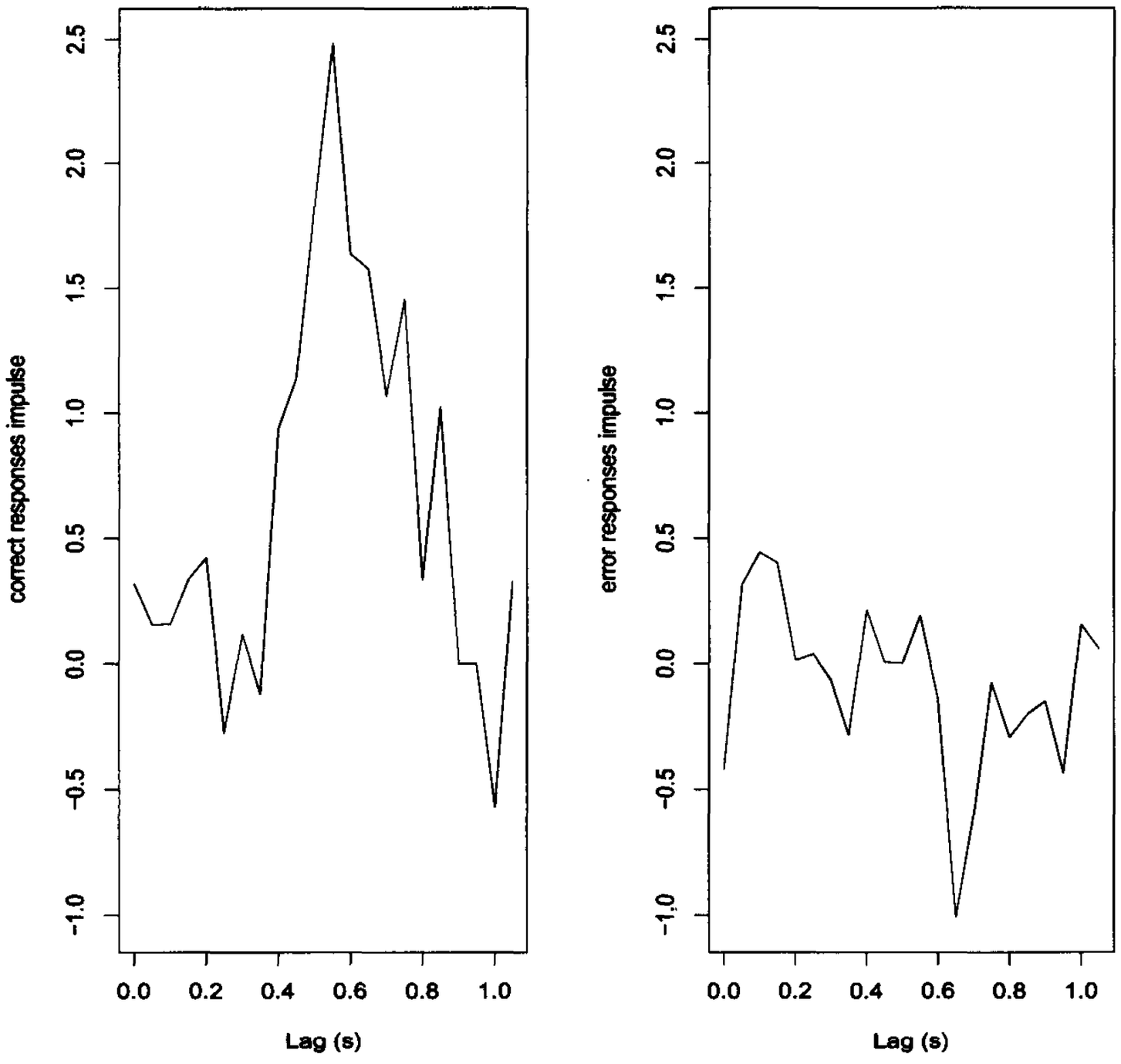


Figure 4.18: Correct (left panel) and error (right panel) response impulses for simulated go-no go RT data with parameters $\mu = .5$, $\sigma = .12$, $d = .1$, $p = .15$, $p_{S_B} = 2$, $p_{S_W} = 2$ and $q = .1$.

is an indication of a larger frequency of responses to the stimulus type at the lag for which the peak occurs. The error impulses have smaller peak values; with a lower frequency the wrong button is pressed.

The majority of the interaction effects are negative, suggesting the presence of inhibition effects among the black flashes. This provides some support for our parametric model, which incorporates thinning among the black flashes. Table 4.10 lists the lags of the interaction coefficients included in the model, as well as the corresponding coefficient value, for each data set. Interaction terms were not found to give an improvement to the fits for the data sets with rates 5.0 and 8.0.

Figure 4.23 displays examples of diagnostic plots, for four of the ten data sets.

Table 4.9: Threshold estimates $\hat{\theta}$ for gonogo RT data.

rate	$\hat{\theta}$
0.4	6.36
0.6	5.55
0.8	5.53
1.0	4.85
1.2	4.80
1.4	4.47
2.0	4.02
4.0	3.22
5.0	2.79
8.0	2.98

The diagnostics suggest that our threshold model fits the data quite well when the flashes are presented with a (combined) rate of at least 1.0. Based on the diagnostic, the threshold models are a poor fit for the data with the slower flash rates of 0.4, 0.6, and 0.8.

The diagnostic plots for the rate 0.4 data (not shown) and 0.6 data (not shown) reveal similar behaviour to the plot for rate 0.8 data. However, the rate 0.6 diagnostic plot indicates a slightly better fit at the upper tail. Furthermore, the rate 0.4 diagnostic suggests that the fitted probabilities tend to underestimate rather than overestimate, as in the rate 0.8 diagnostic.

The rate 1.0 and 1.2 diagnostic plots (not shown) exhibit the same behaviour as the rate 1.4 diagnostic plot, while the rate 4.0 diagnostic (not shown) is similar to the rate 2.0 diagnostic, and the rate 8.0 diagnostic (not shown) resembles the rate 5.0 diagnostic. Each of these diagnostics suggest that a threshold model fits the data quite well.

4.10 Discussion

In this chapter we introduced a parametric model and a threshold model for go-no go RT data, and fit each model to both real and simulated data. Each model

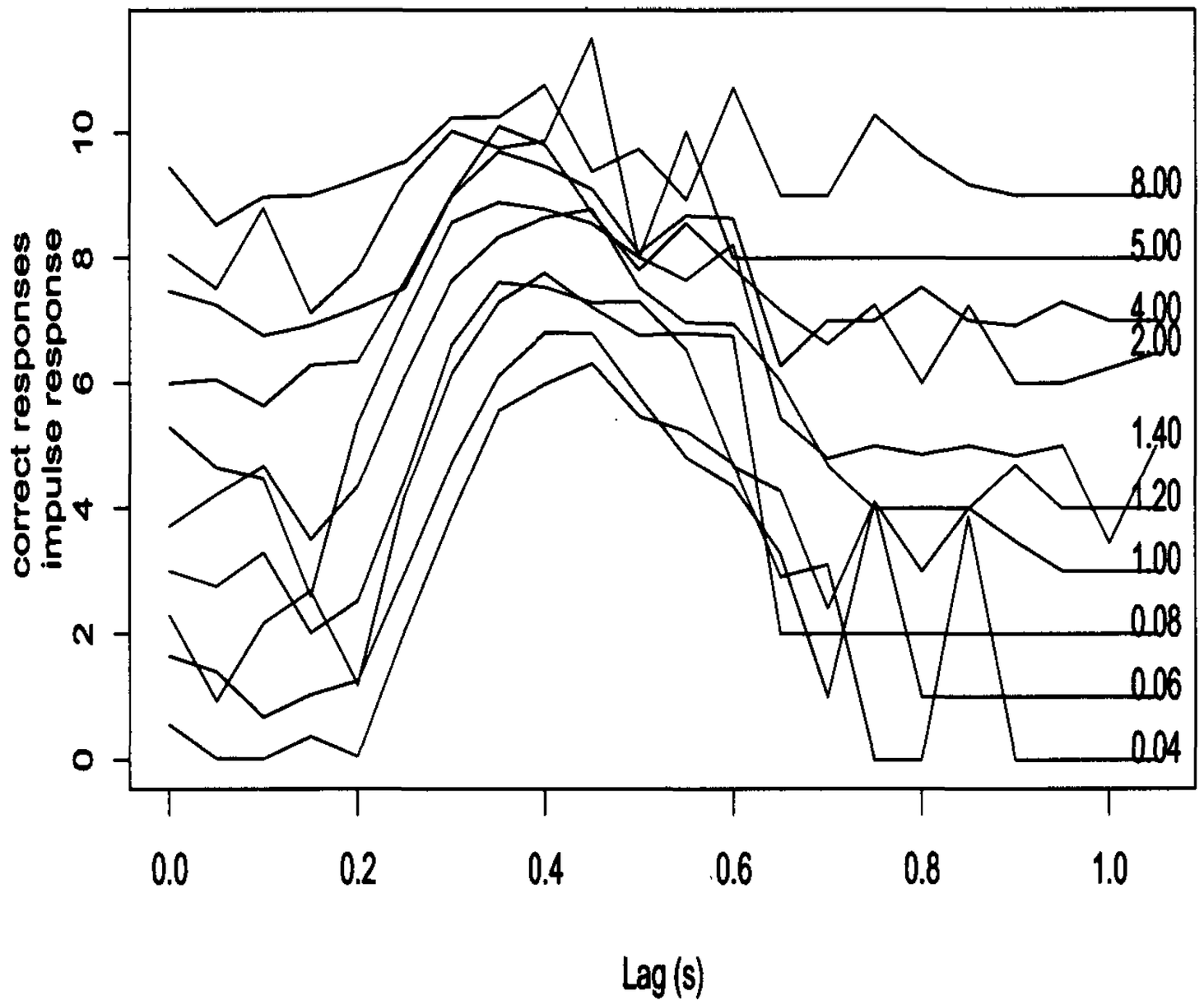


Figure 4.19: Plots of the black stimulus linear filters for each of the 10 go-no go RT data sets. Curves are labelled by the stimulus rate (s). Each curve is shifted upwards by one unit.

is studied for a different reason. The parametric model is of interest for providing a basis for understanding the behavior of nonparametric intensity estimates, while the parameters in a threshold model have direct biological interpretations regarding inferences on the eye-brain-hand system.

Certain point process intensity functions were studied in order to fit the parametric model, and we derived intensity expressions under our model. We also discussed nonparametric intensity estimation, including optimal bandwidth selection. The parametric model was based on the assumption that the reaction times from both black and white flashes are independent and identically distributed as $N(\mu, \sigma)$ random variables. We developed parameter estimation methods under these assump-

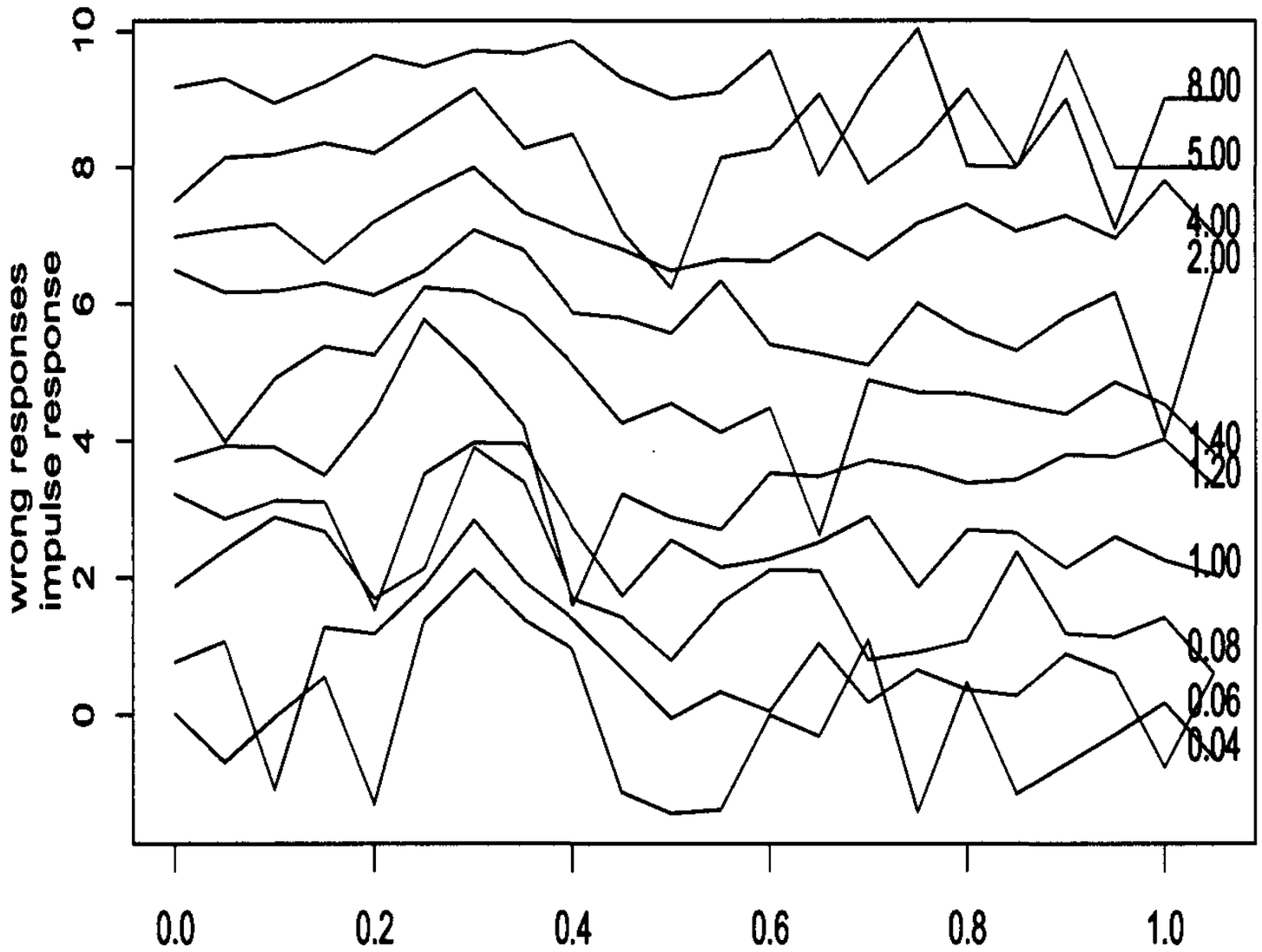


Figure 4.20: Plots of the white stimulus linear filters for each of the 10 go-no go RT data sets. Curves are labelled by the stimulus rate (s). Each curve is shifted upwards by one unit.

tions, and illustrated the methods on simulated data. Finally, we fit our model to ten go-no go RT data sets by applying our estimation methods. For the majority of the data sets the assumption of identical distributions for reaction times from black and white flashes appears to be somewhat unrealistic. However, similar behavior is observed in our simulated data.

Fitting threshold models allowed us to obtain estimates of the threshold, as well as to estimate certain intensities relating the stimuli to the responses. Threshold estimates for the real data sets with different flash rates suggest that as the flash rate increases, a lower rod current is required for a response to occur. There also appears to be evidence of nonlinear inhibition among black flashes.

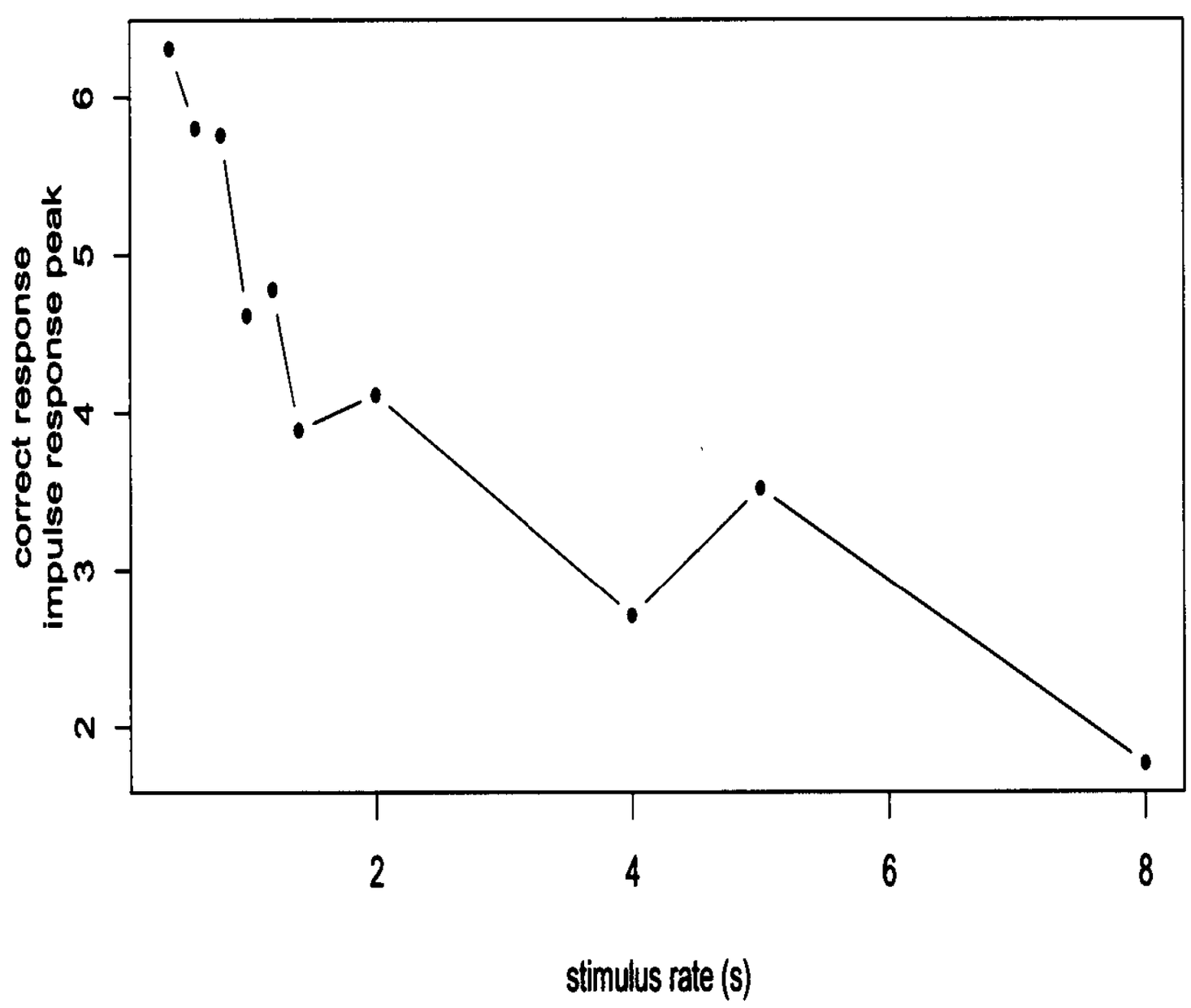


Figure 4.21: Plot of the correct response impulse peak height as a function of stimulus rate.

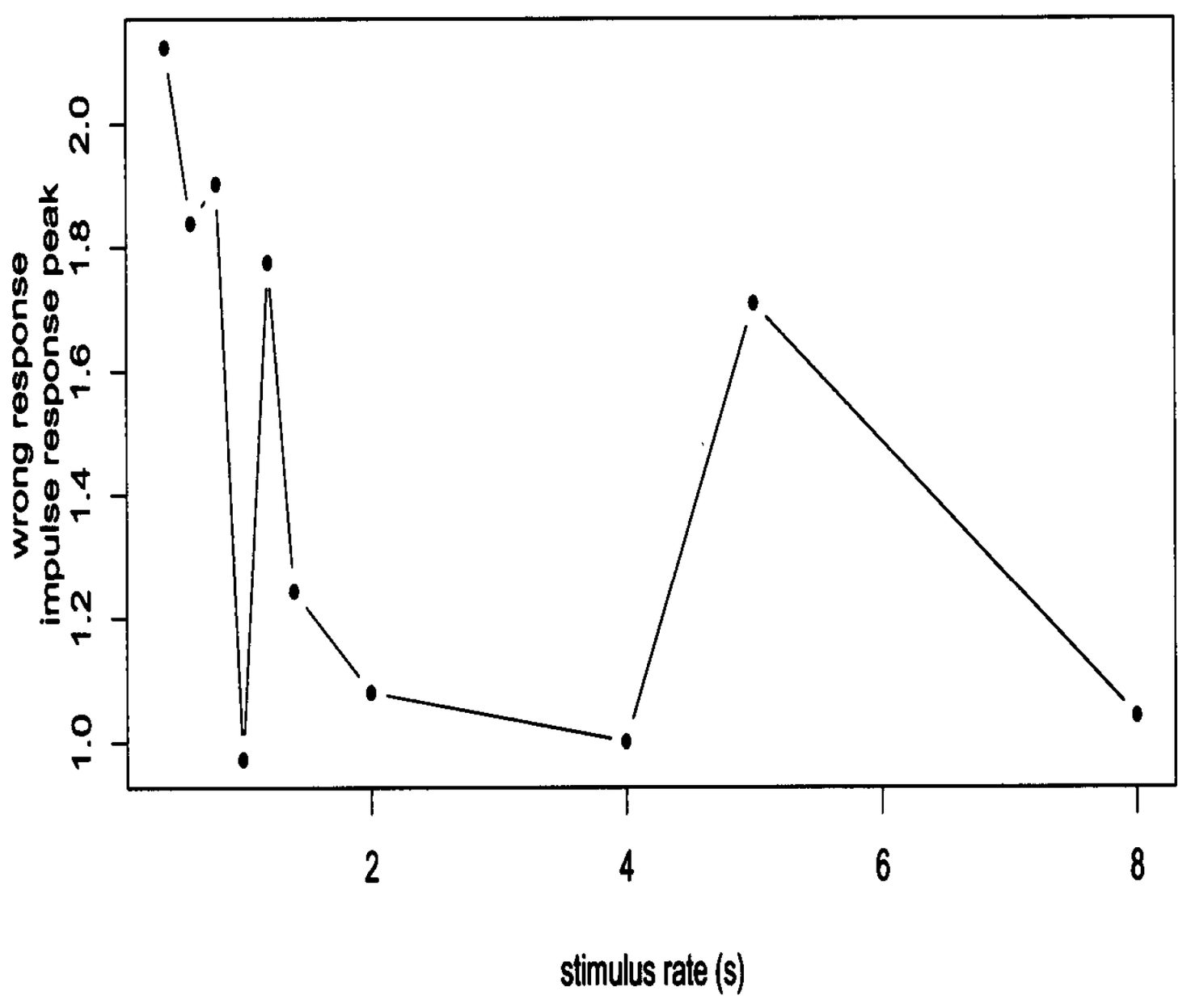


Figure 4.22: Plot of the error impulse response peak height as a function of stimulus rate.

Table 4.10: Black Flash Interactions for go-no go RT data.

rate					
0.4	Lag	.25			
	Coefficient	-2.73			
0.6	Lag	.25	.30	.35	
	Coefficient	-3.13	-4.27	-5.00	
0.8	Lag	.25	.30	.35	
	Coefficient	-2.01	-3.59	-4.86	
1.0	Lag	.3			
	Coefficient	-3.32			
1.2	Lag	.15	.25	.30	.35
	Coefficient	2.32	-1.68	-2.85	-3.98
1.4	Lag	.15	.30		
	Coefficient	2.37	-3.45		
2.0	Lag	.25	.30	.35	.40
	Coefficient	-1.32	-2.79	-2.94	-4.23
4.0	Lag	.10	.35		
	Coefficient	1.06	-1.94		

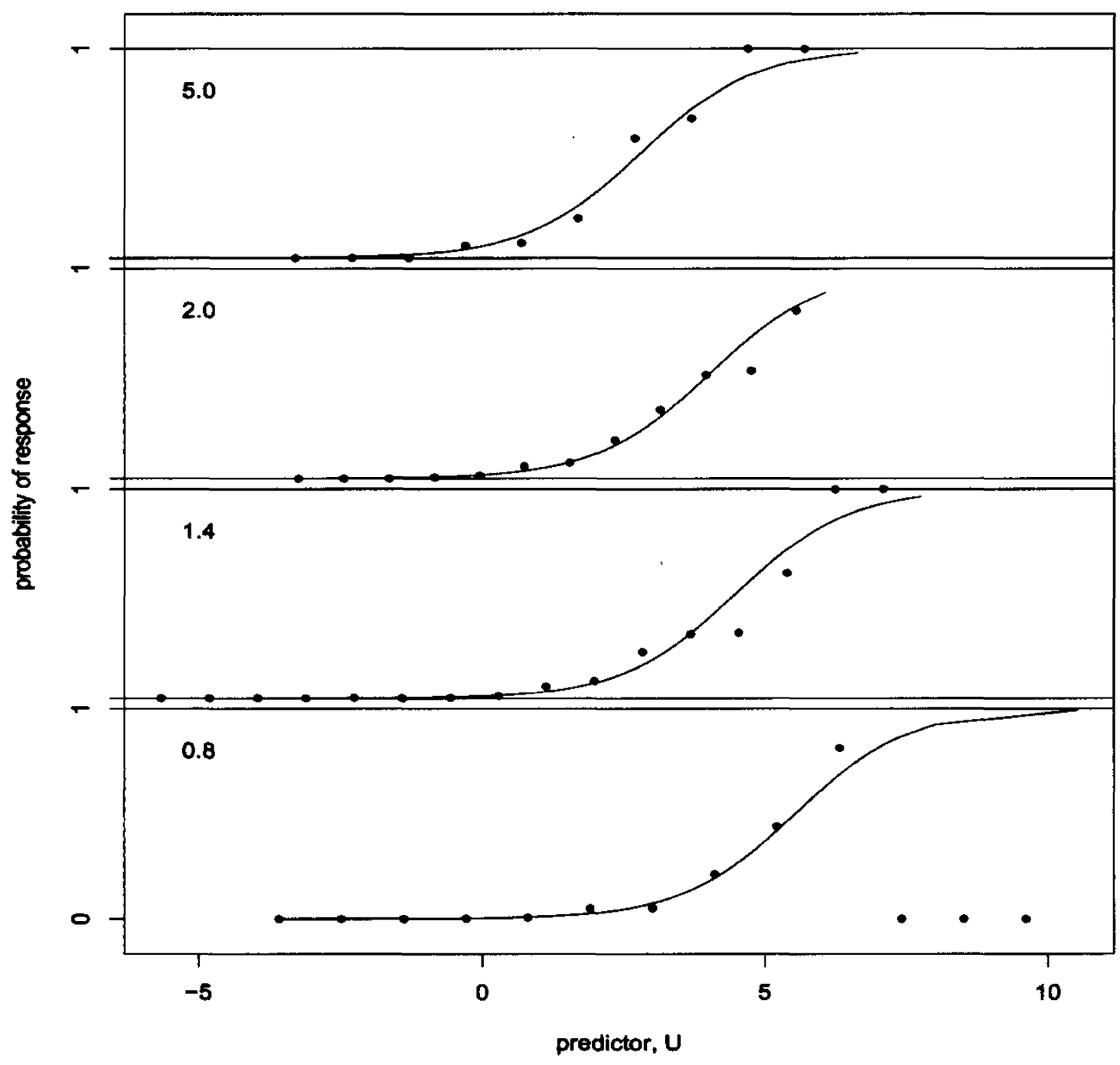


Figure 4.23: Examples of diagnostic plots for the go-no go RT data sets with stimulus rates 0.8,1.4,2.0 and 5.0, as indicated in the top left corner of each plot. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

Chapter 5

A Parametric Model for the Choice RT Experiment

In the choice RT experiment, there are two types of stimuli, (black or white flashes) and two types of responses ('black' or 'white' button presses). The subject may correctly press the 'white' button in response to a white flash, or may press it as one of two possible errors. First, the incorrect 'white' response may be due to misperceiving a black flash. Another possibility is pressing the 'white' button when neither flash type is presented. The latter error type is due to internal noise within the eye-brain-hand system that causes the observer to mistakenly perceive a white stimulus. Similar kinds of errors occur for black responses.

The choice RT experiment is more closely related to 'real' activities than either the simple RT or go-no go RT experiments. In a choice RT experiment the subject has to decide when a flash occurs and the type of flash presented, as well as press one of two buttons corresponding to the flash type. Although not as complicated, the combination of activities required in a choice RT experiment is similar to responding to traffic lights when driving a car. Considering red and green lights, a driver needs to decide which type of light is present, and choose the correct action depending on the light type - press down on brake pedal for red, and acceleration pedal for green.

The goal of this chapter is to introduce a parametric model for choice RT data, and use it in conjunction with nonparametric estimation, to make inferences about

the eye-brain-hand system.

In Section 5.1 we introduce a parametric model for a choice RT experiment and describe the point process operations involved. We outline how to simulate data from this model in Section 5.2. Next, in Section 5.3 we discuss some point process intensity functions for this model. In the subsequent section we provide the derivations of these intensities. We describe nonparametric intensity estimation and discuss bandwidth selection in Section 5.5. In Section 5.6 we develop parameter estimation methods, while in Section 5.7 we illustrate our methods for simulated data. Finally, in section 5.8, we fit our model to nine data sets from choice RT experiments with different flash rates. A discussion and conclusions follow in Section 5.9.

5.1 The Proposed Model

In a choice reaction time experimental run, black and white flashes are presented to a subject as independent homogeneous Poisson processes with rates p_{S_B} and p_{S_W} (per second), respectively. Thus, the entire flash process, which is the superposition of the black and white flash processes, has rate $p_{S_B} + p_{S_W}$.

We denote the number of white flashes in the time interval X by $S_W(X)$, and the set of times of the white flashes is denoted by $S_W = \{S_{W,1}, S_{W,2}, \ldots\}$. Analogous notation is used for the black flash times, S_B, as well as the subject's white response times to both types of flashes, R_W, and the black response times to both stimulus types, R_B. Note that S_B and S_W are taken to be independent stationary Poisson processes.

As in Chapter 4, we postulate internal noise. We associate a noise process with each of the black and white flash processes, and model these noise processes as independent Poisson processes S_{N_B} and S_{N_W}, with respective rates p_{N_B} and p_{N_W}. Each noise process is superposed onto the corresponding flash process, so that the superposed processes are $S'_B = S_B \bigcup S_{N_B}$ and $S'_W = S_W \bigcup S_{N_W}$.

The deletion probability for the thinning of white flashes is given by

$$P(S'_{W,j} \text{ deleted}) = \begin{cases} 1, & S'_{W,j+1} - S'_{W,j} < d \\ p, & \text{otherwise} \end{cases} . \tag{5.1.1}$$

The deletion probability for the thinning of the black flashes is expressed as in (5.1.1), but with all instances of W replaced by B. We denote the thinned white and black flash processes by S''_W and S''_B, respectively.

Each response is of the form $R_i = S''_i + V_i$, where $\{V_i\}$ is a sequence of independent and identically distributed random variables with common density $f_V(v)$. The random variables V_i correspond to the amounts of time required for the brain and hand to respond to the flash; these are the particular reaction times for each of the events.

The responses are related to the flash processes as follows, where q is the probability of responding incorrectly to a flash, and we make the simplifying assumption that the delays V_i for black and for white flashes are identically distributed

$$S''_{W,j} + V_j \rightarrow \begin{cases} R_{W,i}, & \text{w.p. } 1-q \\ R_{B,i}, & \text{w.p. } q \end{cases} ,$$

$$S''_{B,j} + V_j \rightarrow \begin{cases} R_{B,i}, & \text{w.p. } 1-q \\ R_{W,i}, & \text{w.p. } q \end{cases} .$$

Response types are completely random in the case where $q = .5$.

We assume that the reaction times from black flashes are independent of the reaction times from white flashes. The parametric intensity functions that we derive in Section 5.4 depend on this assumption.

We now set out all of the assumptions that we make in this chapter:

Assumptions

1. Reaction times resulting from black (white) flashes are independent random variables.

2. Reaction times from black flashes are independent of those from white flashes.

3. Reaction times from white flashes have the same distribution as those from black flashes.

4. S_B and S_W are independent stationary Poisson processes.

5. All reaction times have a $N(\mu, \sigma)$ distribution. However, the derivations and estimation methods are also valid for alternative distributions.

5.2 Simulation of Data from the Model

The flash process is simulated as it was for the go-no go experiment. A noise process is simulated in a manner similar to that used for the go-no experiment. However, the noise process rate is now $p_{N_B} + p_{N_W}$, and each noise event is assigned an identification label of '0' or '1'. The probability that a noise event is associated with the black flash process is $\frac{p_{N_B}}{p_{N_B}+p_{N_W}}$, and the probability that it is associated with the white flashes is $\frac{p_{N_W}}{p_{N_B}+p_{N_W}}$.

The data set, resulting from Algorithm 5.2.1, will be a list consisting of four vectors, not necessarily the same length, consisting of stimulus times and response times, as well as identification labels for the times. That is, each black stimulus (response) time has a label of '0', while a label of '1' identifies those that are white.

Algorithm 5.2.1. *The following steps can be used to generate choice RT data consisting of N_S flashes with black flash rate p_{S_B}, white flash rate p_{S_W}, black noise rate p_{N_B}, white noise rate p_{N_W}, thinning parameter d, thinning probability p, and $N(\mu, \sigma)$ distributed reaction times.*

1. *Generate stimulus sequences consisting of a total of N_S black and white stimuli.*

   (a) *Generate N_S random exponential variates with mean $\frac{1}{p_{S_B}+p_{S_W}}$.*

   (b) *Compute the N_S cumulative sums of the exponential variates. Assign these times to S.*

(c) Set

$$p_0 \leftarrow \frac{p_{S_B}}{p_{S_B} + p_{S_W}}.$$

(d) Assign the stimulus types. For $i = 1, 2, \ldots, N_S$, *set*

$$ID_S[i] \leftarrow \begin{cases} 0 & \textit{with probability } p_0 \\ 1 & \textit{with probability } 1 - p_0 \end{cases}$$

(e) Set $S_B \leftarrow \{S[i] : ID_S[i] = 0\}$ *and* $S_W \leftarrow \{S[i] : ID_S[i] = 1\}$.

2. Add noise. If $p_{N_B} + p_{N_W} > 0$, *then*

(a) Generate a Poisson $(p_{N_B} + p_{N_W})T$ *random variate and assign to* N_N.

(b) If $N_N = 0$, *go to Step 3.*

(c) Assign $N_S \leftarrow N_S + N_N$.

(d) Generate N_N *random uniform variates on* $[0, T]$. *Assign these values to* N.

(e) Set

$$p_1 \leftarrow \frac{p_{N_B}}{p_{N_B} + p_{N_W}}.$$

(f) Assign the stimulus types to the noise. For $i = 1, 2, \ldots, N_N$, *set*

$$ID_{S_N}[i] \leftarrow \begin{cases} 0 & \textit{with probability } p_1 \\ 1 & \textit{with probability } 1 - p_1 \end{cases}$$

(g) Set $S' \leftarrow S \cup N$ *and* $ID_{S'_N} \leftarrow ID_S \cup ID_{S_N}$.

(h) Sort S' *and* $ID_{S'_N}$ *in ascending order of* S'.

3. Thinning and translating.

(a) Set $N_R \leftarrow 0$.

(b) For $i = 2, \ldots, N_{S_B}$, *if* $(S'_B[i] - S'_B[i-1] > d)$, *then do the following with probability* $1 - p$:

i. Assign $N_R \leftarrow N_R + 1$.

ii. *Generate a* $N(\mu, \sigma^2)$ *random variable* V.

iii. *Assign* $R[N_R] \leftarrow S'_B[i-1] + V$

iv. *Set*

$$ID_R[N_R] \leftarrow \begin{cases} 0 & \textit{with probability } 1-q \textit{ (correct response)} \\ 1 & \textit{with probability } q \textit{ (incorrect response)} \end{cases}$$

(c) For $i = 2, \ldots, N_{S_W}$, *if* $(S'_W[i] - S'_W[i-1] > d)$, *then do the following with probability* $1-p$*:*

i. *Assign* $N_R \leftarrow N_R + 1$.

ii. *Generate a* $N(\mu, \sigma^2)$ *random variable* V.

iii. *Assign* $R[N_R] \leftarrow S'_W[i-1] + V$

iv. *Set*

$$ID_R[N_R] \leftarrow \begin{cases} 1 & \textit{with probability } 1-q \textit{ (correct response)} \\ 0 & \textit{with probability } q \textit{ (incorrect response)} \end{cases}$$

4. Sort R *and* ID_R *in ascending order of* R.

5. Return S, ID_S, R, ID_R.

We used the programming language Fortran (Lahey Computer Systems, Inc. (1999)) to implement our algorithm.

5.3 Point Process Intensity Functions

We denote the white response rate, or first order intensity of the R_W process by p_{R_W}. Analogously, we denote the black response rate by p_{R_B}.

There are four stimulus-response second order intensities corresponding to each possible combination of black and white flashes with black and white responses, which we denote $p_{S_W R_W}(u)$, $p_{S_W R_B}(u)$, $p_{S_B R_W}(u)$, and $p_{S_B R_B}(u)$. The intensities $p_{S_W R_W}(u)$ and $p_{S_B R_B}(u)$ correspond to intensities of a response following a flash of the same type at a lag of u time units, while $p_{S_B R_W}(u)$ and $p_{S_W R_B}(u)$ coincide with intensities of a flash followed by the wrong type of response at a lag of u time units.

Corresponding to each possible combination of two responses, there are four response-response second order intensities, which we denote $p_{R_W R_W}(u)$, $p_{R_W R_B}(u)$, $p_{R_B R_W}(u)$, and $p_{R_B R_B}(u)$. The intensities $p_{R_s R_r}(u)$ satisfy

$$E[R_s(X)R_r(Y)] = \begin{cases} \int_X \int_Y p_{R_s R_r}(y-x)\,dx\,dy, & r \neq s \\ \int_X \int_Y p_{R_r R_r}(y-x)\,dx\,dy + p_{R_r}|X \cap Y|_L, & r = s, \end{cases}$$

where r and s take on values in $\{B, W\}$. Since the above expressions remain the same in the event that the time intervals X and Y are shifted by h units, we have that the response processes are second order stationary processes.

Under our model, the point process (S_B, S_W, R_B, R_W) is Brillinger mixing. This can be shown in a manner similar to the argument given in Section 4.3 to show that, under the model assumptions, (S_B, S_W, R) is Brillinger mixing.

For our parametric model, the response rates are given by

$$p_{R_W} = (1-p)[q(p_{S_B} + p_{N_B})e^{-d(p_{S_B}+p_{N_B})} + (1-q)(p_{S_W} + p_{N_W})e^{-d(p_{S_W}+p_{N_W})}], \quad (5.3.1)$$

and

$$p_{R_B} = (1-p)[q(p_{S_W} + p_{N_W})e^{-d(p_{S_W}+p_{N_W})} + (1-q)(p_{S_B} + p_{N_B})e^{-d(p_{S_B}+p_{N_B})}]. \quad (5.3.2)$$

The second order stimulus-response intensities are given by:

$$\begin{aligned} p_{S_W R_W}(u) = \quad & (1-p)p_{S_W}[q(p_{S_B} + p_{N_B})e^{-d(p_{S_B}+p_{N_B})} \\ & + (1-q)e^{-d(p_{S_W}+p_{N_W})}\{(p_{S_W} + p_{N_W})P(V \notin (u, u+d)) + f(u)\}] \end{aligned} \quad (5.3.3)$$

$$\begin{aligned} p_{S_W R_B}(u) = \quad & (1-p)p_{S_W}[(1-q)(p_{S_B} + p_{N_B})e^{-d(p_{S_B}+p_{N_B})} \\ & + qe^{-d(p_{S_W}+p_{N_W})}\{(p_{S_W} + p_{N_W})P(V \notin (u, u+d)) + f(u)\}] \end{aligned} \quad (5.3.4)$$

$$\begin{aligned} p_{S_B R_B}(u) = \quad & (1-p)p_{S_B}[q(p_{S_W} + p_{N_W})e^{-d(p_{S_W}+p_{N_W})} \\ & + (1-q)e^{-d(p_{S_B}+p_{N_B})}\{(p_{S_B} + p_{N_B})P(V \notin (u, u+d)) + f(u)\}] \end{aligned} \quad (5.3.5)$$

$$\begin{aligned} p_{S_B R_W}(u) = \quad & (1-p)p_{S_B}[(1-q)(p_{S_W} + p_{N_W})e^{-d(p_{S_W}+p_{N_W})} \\ & + qe^{-d(p_{S_B}+p_{N_B})}\{(p_{S_B} + p_{N_B})P(V \notin (u, u+d)) + f(u)\}], \end{aligned} \quad (5.3.6)$$

and the response-response intensities are expressed as follows:

$$\begin{aligned} p_{R_W R_W}(u) = \quad & (1-p)^2[P(|V_2 - V_1 - u| > d)\{q^2(p_{S_B} + p_{N_B})^2 e^{-2d(p_{S_B}+p_{N_B})} \\ & +(1-q)^2(p_{S_W} + p_{N_W})^2 e^{-2d(p_{S_W}+p_{N_W})}\} \\ & +2q(1-q)(p_{S_B} + p_{N_B})(p_{S_W} + p_{N_W}) e^{-d(p_{S_B}+p_{N_B}+p_{S_W}+p_{N_W})}] \quad (5.3.7) \end{aligned}$$

$$\begin{aligned} p_{R_B R_B}(u) = \quad & (1-p)^2[P(|V_2 - V_1 - u| > d)\{q^2(p_{S_W} + p_{N_W})^2 e^{-2d(p_{S_W}+p_{N_W})} \\ & +(1-q)^2(p_{S_B} + p_{N_B})^2 e^{-2d(p_{S_B}+p_{N_B})}\} \\ & +2q(1-q)(p_{S_W} + p_{N_W})(p_{S_B} + p_{N_B}) e^{-d(p_{S_W}+p_{N_W}+p_{S_B}+p_{N_B})}] \quad (5.3.8) \end{aligned}$$

$$\begin{aligned} p_{R_W R_B}(u) = \quad & (1-p)^2[P(|V_2 - V_1 - u| > d)\{q(1-q)(p_{S_B} + p_{N_B})^2 e^{-2d(p_{S_B}+p_{N_B})} \\ & +q(1-q)(p_{S_W} + p_{N_W})^2 e^{-2d(p_{S_W}+p_{N_W})}\} \\ & +q^2(p_{S_B} + p_{N_B})(p_{S_W} + p_{N_W}) e^{-d(p_{S_B}+p_{N_B}+p_{S_W}+p_{N_W})} \\ & +(1-q)^2(p_{S_B} + p_{N_B})(p_{S_W} + p_{N_W}) e^{-d(p_{S_B}+p_{N_B}+p_{S_W}+p_{S_B})}], \quad (5.3.9) \end{aligned}$$

where V, V_1, and V_2 are independent random variables having common density $f_V(v)$. By definition $p_{R_B R_W}(u) = p_{R_W R_B}(u)$.

Graphs of the four stimulus-response (S-R) second order intensities with three values of q at two values of d are plotted in Figures 5.1 and 5.2. These intensities have a peak at the mode of $f(u)$, which increases in height as the stimulus rate increases. The effect of the q parameter as it approaches zero is to flatten the black-white and white-black S-R intensity curves; when $q = 0$ these intensities are constant. This behavior is expected since when q is near zero there is a very low probability of an error response type, so there is a low intensity of black responses following white flashes, and vice versa. As q increases to 0.5 the four curves have similar peak heights. In the case of $q = .5$, after a flash there is an equal probability of either response type, thus the intensities of each kind of response following either stimulus are similar. Thus, information about the magnitude of the error probability q is contained in the white-black and black-white S-R intensities. As the d parameter increases the curve to the left of the peak is depressed. Note that the d parameter has the same effect on the stimulus-response second order intensity for the simple RT model described in Braun et al. (2003).

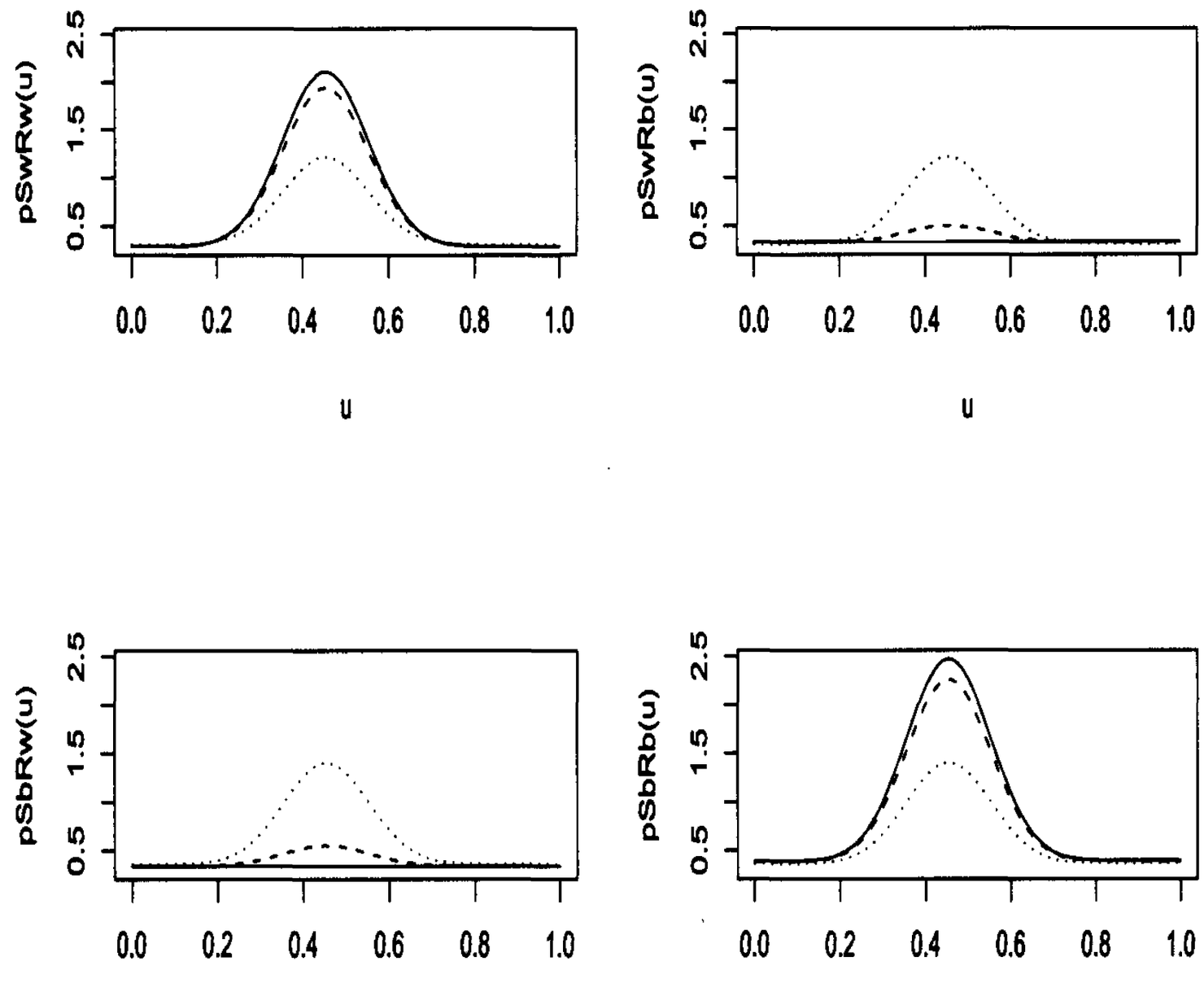


Figure 5.1: The second-order stimulus-response intensity functions with $p_{S_W} = .6$, $p_{S_B} = .7$, μ=.45, σ=.1, p=.15, $p_{N_B} = p_{N_W} = 0$, $d = .1$, and $q = 0$ (solid line), $q = .1$ (dashed line), and $q = .5$ (dotted line).

Figure 5.3 displays plots of the second order response-response (R-R) intensities for three values of d at two values of q. Since the $p_{R_W R_B}(u)$ and $p_{R_B R_W}(u)$ functions are identical, graphs of $p_{R_B R_W}(u)$ are excluded. When $d = 0$ the intensities are constant. Otherwise, there is a trough centered at zero, which widens as d increases. Thus, information regarding the magnitude of the nonlinear inhibition parameter d is contained in the white-white and black-black R-R intensities. In addition, as q approaches zero the black-white and white-black R-R intensity curves become more narrow and lose some depth.

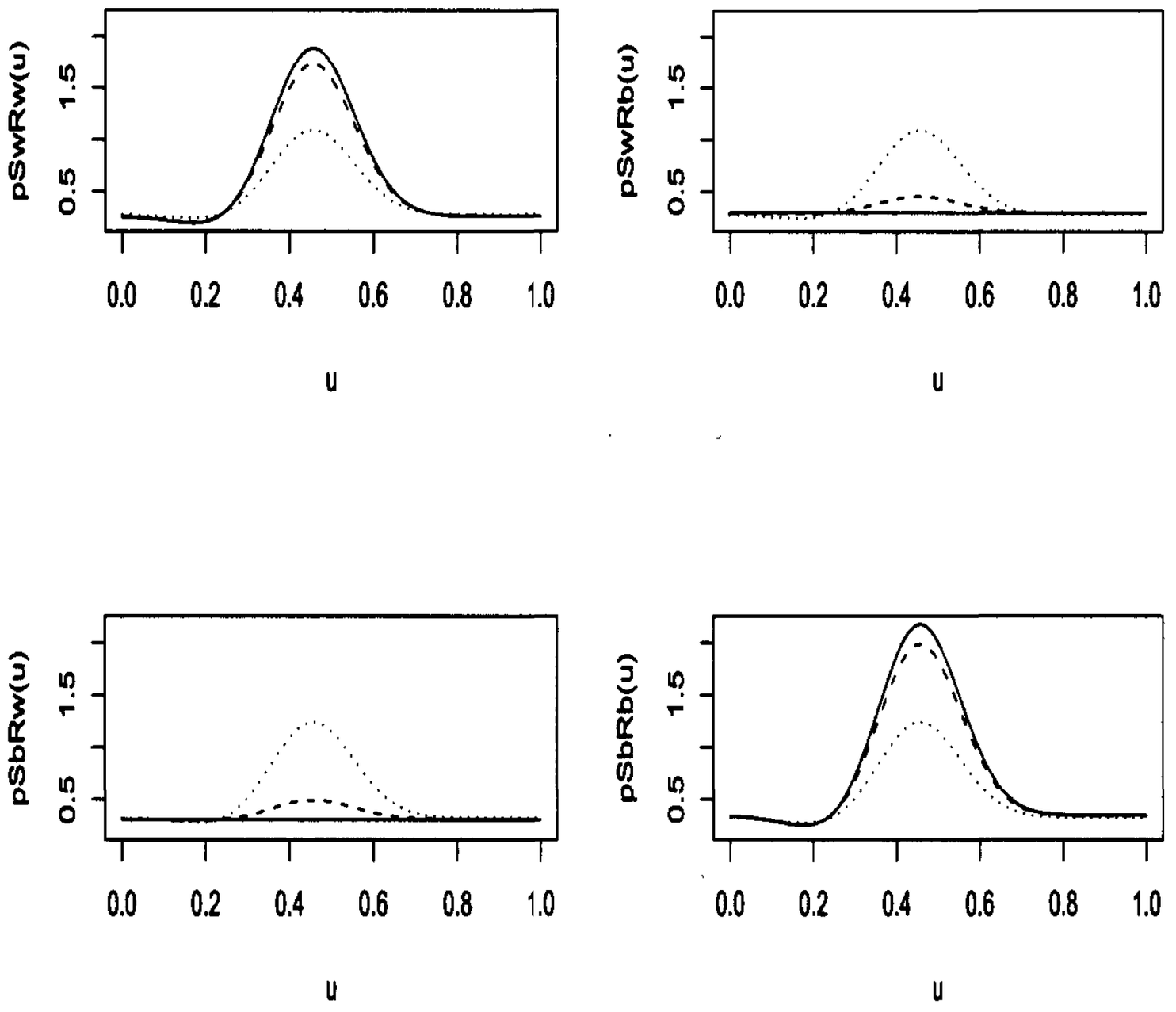


Figure 5.2: The second-order stimulus-response intensity functions with $p_{S_W} = .6$, $p_{S_B} = .7$, μ=.45, σ=.1, p=.15, $p_{N_B} = p_{N_W} = 0$, $d = .25$, and $q = 0$ (solid line), $q = .1$ (dashed line), and $q = .5$ (dotted line).

5.4 Derivations of the Intensity Functions

5.4.1 First-Order Response Intensities

We first derive the expression for the first order intensity of the R_W process. We have

$$p_{R_W} = \lim_{h \to 0} h^{-1} E[R_W(u, u+h)]. \tag{5.4.1}$$

As a result of the translation mechanism in the model, we have

$$R_W(u, u+h) = \sum_i 1_{\{S''_{B,i} \in (u-V_i, u-V_i+h)\}} A_i + \sum_i 1_{\{S''_{W,i} \in (u-V_i, u-V_i+h)\}} B_i,$$

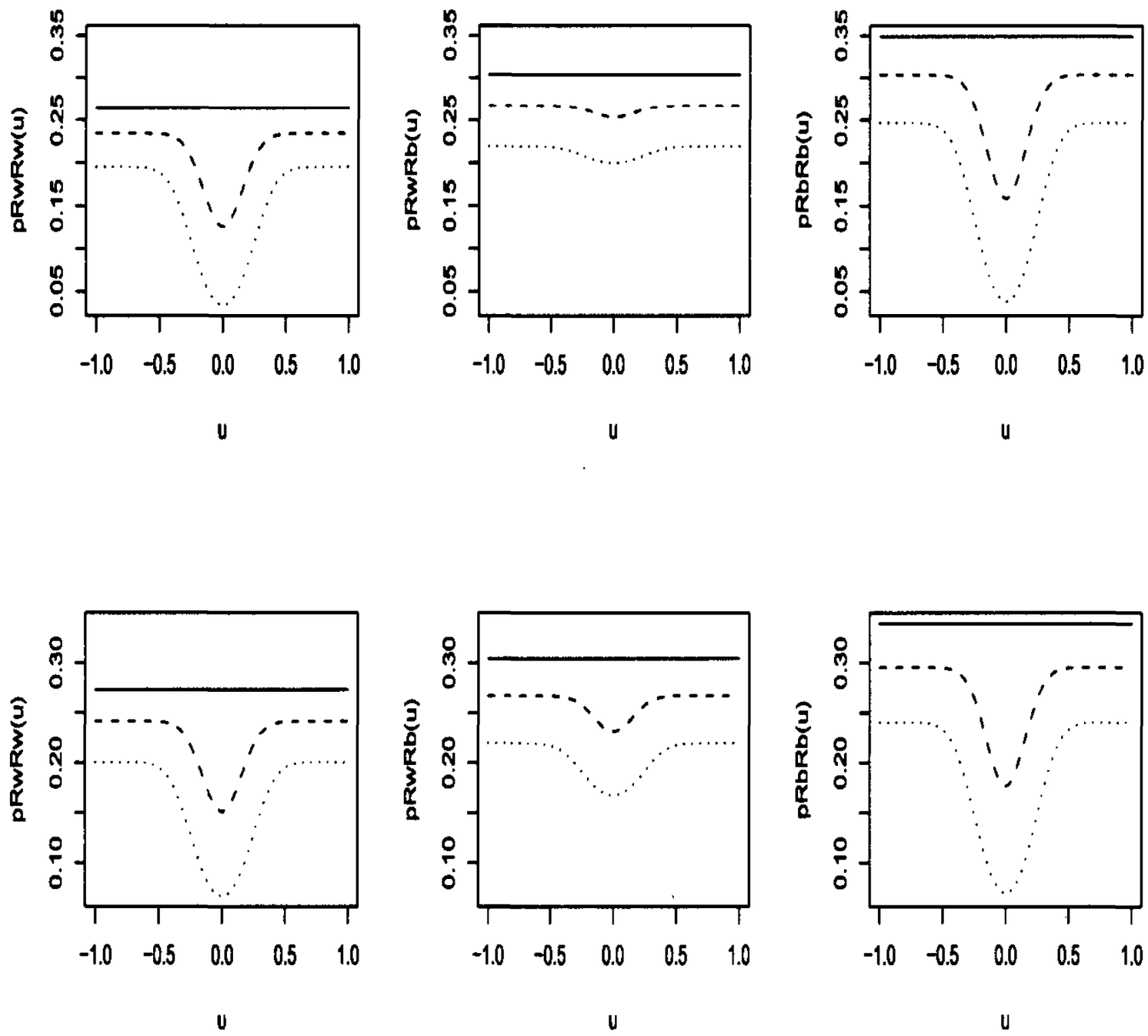


Figure 5.3: The second-order response-response intensity functions with $p_{S_W} = .6$, $p_{S_B} = .7$, μ=.45,σ=.1, p=.15, $p_{N_B} = p_{N_W} = 0$, $d = 0$ (solid line), $d = .1$ (dashed line), and $d = .25$ (dotted line), and $q = .05$ (top panel), $q = .15$ (bottom panel)

where

$$A_i = 1_{\{\text{white response}|\text{black flash at}S_i'\}}$$

$$B_i = 1_{\{\text{white response}|\text{white flash at}S_i'\}},$$

$S_i' = (S_B'' \bigcup S_W'')_i$, and the V_i represent independent random variables from a population having density $f(v)$. As a result of the random splitting in the model, the probability of a white response given a black flash is q, and the white response prob-

ability conditional on a white flash is $1-q$. Thus,

$$\begin{aligned} & E[R_W(u,u+h)] \\ &= qE\left[\sum_i 1_{\{S''_{B,i}\in(u-V_i,u-V_i+h)\}}\right] + (1-q)E\left[\sum_i 1_{\{S''_{W,i}\in(u-V_i,u-V_i+h)\}}\right]. \end{aligned} \tag{5.4.2}$$

The first term in (5.4.2) corresponds to the rate of the white response process formed from black flashes, while the second term coincides with the rate of the white response process formed from white flashes.

When there is no non-linear inhibition (i.e. $d=0$) flashes are deleted with probability p independently of the other flashes, so we have

$$E[S''_W(dx_1)] = E[S'_W(dx_1)1_{\{U_{y_1-v}>p\}}] = (1-p)(p_{S_W}+p_{N_W})\,dx_1,$$

where U_{y_1-v} is a uniform random variable that is independent of the flash and response processes. Thus the thinned white flashes are a Poisson process with rate $p_{S''_W} = (p_{S_W}+p_{N_W})(1-p)$. Similarly, the rate of the thinned black flash process is $p_{S''_B} = (p_{S_B}+p_{N_B})(1-p)$.

When $d>0$, a flash is deleted with probability p when there is a flash of the same type within d time units after it, so we have

$$\begin{aligned} E[S''_W(d(x_1))] &= E[S'_W(d(x_1))1_{\{U_{x_1}>p\}}1_{\{S'_W(x_1,x_1+d)=0\}}] \\ &= (1-p)(p_{S_W}+p_{N_W})e^{-d(p_{S_W}+p_{N_W})}\,dx_1. \end{aligned} \tag{5.4.3}$$

Similar expressions can be obtained for the black flashes, by interchanging the 'B' and 'W' subscripts.

Thus, we have

$$
\begin{aligned}
& E[R_W(u,u+h)] \\
&= qE[\sum_i 1_{\{S''_{B,i}\in(u-V_i,u-V_i+h)\}}] + (1-q)E[\sum_i 1_{\{S''_{W,i}\in(u-V_i,u-V_i+h)\}}] \\
&= qE[E[\sum_i 1_{\{S''_{B,i}\in(u-V_i,u-V_i+h)\}}|S''_B]] \\
&\quad +(1-q)E[E[\sum_i 1_{\{S''_{W,i}\in(u-V_i,u-V_i+h)\}}|S''_W]] \\
&= qE[E[E[\sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}|V,S''_B]] \\
&\quad +(1-q)E[E[E[\sum_i 1_{\{S''_{W,i}\in(u-v,u-v+h)\}}|V,S''_W]]] \\
&= qE[E[\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}} f(v)\,dv|S''_B]] \\
&\quad +(1-q)E[E[\sum_i \int 1_{\{S''_{W,i}\in(u-v,u-v+h)\}} f(v)\,dv|S''_W]] \\
&= qE[\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}} f(v)\,dv] \qquad (5.4.4) \\
&\quad +(1-q)E[\sum_i \int 1_{\{S''_{W,i}\in(u-v,u-v+h)\}} f(v)\,dv] \\
&= qE[\int \sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}} f(v)\,dv] \\
&\quad +(1-q)E[\int \sum_i 1_{\{S''_{W,i}\in(u-v,u-v+h)\}} f(v)\,dv] \qquad (5.4.5) \\
&= qE[\int S''_B(u-v,u-v+h) f(v)\,dv] \\
&\quad +(1-q)E[\int S''_W(u-v,u-v+h) f(v)\,dv] \qquad (5.4.6)
\end{aligned}
$$

where (5.4.4) is obtained by using the fact that the V_i are i.i.d. with density $f(\cdot)$. By application of the Tonelli-Fubini theorem (see, Jacod and Protter (1991), pp. 63-64), the integral and sum are reversed to get (5.4.5).

Since S''_W, and S''_B are counting measures, and for a counting measure $A(\cdot)$ we have $A(0,h) = \int_0^h A(dx_1)$ (see Daley and Vere-Jones (2003)), (5.4.6) may be expressed as

$$qE[\int \int_u^{u+h} S''_B(d(y_1-v)) f(v)\,dv] + (1-q)E[\int \int_u^{u+h} S''_W(d(y_1-v)) f(v)\,dv].$$

By noting that the integrand is non-negative the order of integration and expectation

may be reversed yielding

$$E[R_W(u,u+h)]$$
$$= q\int_u^{u+h}\int E[S''_B(d(y_1-v))]f(v)\,dv + (1-q)\int_u^{u+h}\int E[S''_W(d(y_1-v))]f(v)\,dv.$$

Using (5.4.3) and the stationarity of S''_B and S''_W it follows that

$$\begin{aligned} & E[R_W(u,u+h)] \\ = \; & q\int_u^{u+h}\int (1-p)(p_{S_B}+p_{N_B})e^{-d(p_{S_B}+p_{N_B})}\,dy_1 f(v)\,dv \\ & +(1-q)\int_u^{u+h}\int (1-p)(p_{S_W}+p_{N_W})e^{-d(p_{S_W}+p_{N_W})}\,dy_1 f(v)\,dv \\ = \; & h(1-p)\int f(v)\,dv[q(p_{S_B}+p_{N_B})e^{-d(p_{S_B}+p_{N_B})} \\ & +(1-q)(p_{S_W}+p_{N_W})e^{-d(p_{S_W}+p_{N_W})}]. \end{aligned} \tag{5.4.7}$$

Substitution of (5.4.7) into (5.4.1), and using the fact that $f(v)$ is a probability density function, we see that the first order intensity for the R_W process is given as in (5.3.1):

$$p_{R_W} = (1-p)[q(p_{S_B}+p_{N_B})e^{-d(p_{S_B}+p_{N_B})} + (1-q)(p_{S_W}+p_{N_W})e^{-d(p_{S_W}+p_{N_W})}].$$

In the same manner, expression (5.3.2) can be derived for p_{R_B}.

5.4.2 Stimulus-Response Second-Order Intensities

We now derive expression (5.3.3) for $p_{S_W R_W}(u)$. A similar argument can be used to obtain expressions (5.3.4), (5.3.5), and (5.3.6) for $p_{S_W R_B}(u)$, $p_{S_B R_B}(u)$ and $p_{S_B R_W}(u)$, respectively.

The intensity $p_{S_W R_W}(u)$ is generally a function of two variables, but it reduces to a function of one variable because of stationarity. In other words, letting $u = y - x$,

$$p_{S_W R_W}(x,y) = \lim_{h\to 0} h^{-2}E[S_W(x,x+h)R_W(y,y+h)]$$

may be expressed as

$$p_{S_W R_W}(u) = \lim_{h\to 0} h^{-2}E[S_W(0,h)R_W(u,u+h)]. \tag{5.4.8}$$

In the following, we assume that the reaction times V_i from the white and black flashes are independently and identically distributed with the same distribution $f(v)$. That is, this result depends on Assumptions 1, 2, and 3.

$$
\begin{aligned}
& E[S_W(0,h)R_W(u,u+h)] \\
= \; & E[S_W(0,h)(qE[\sum_i 1_{\{S''_{B,i}\in(u-V_i,u-V_i+h)\}}|S''_W,S''_B] \\
& +(1-q)E[\sum_i 1_{\{S''_{W,i}\in(u-V_i,u-V_i+h)\}}|S''_W,S''_B])] \\
= \; & E[S_W(0,h)(qE[E[\sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}|V,S''_B]] \\
& +(1-q)E[E[\sum_i 1_{\{S''_{W,i}\in(u-v,u-v+h)\}}|V,S''_W]])] \qquad (5.4.9) \\
= \; & E[S_W(0,h)(qE[\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}f(v)\,dv|S''_B] \\
& +(1-q)E[\sum_i \int 1_{\{S''_{W,i}\in(u-v,u-v+h)\}}f(v)\,dv|S''_W])] \\
= \; & E[S_W(0,h)(q\sum_i \int 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}f(v)\,dv \\
& +(1-q)\sum_i \int 1_{\{S''_{W,i}\in(u-v,u-v+h)\}}f(v)\,dv)] \\
= \; & E[S_W(0,h)(q\int \sum_i 1_{\{S''_{B,i}\in(u-v,u-v+h)\}}f(v)\,dv \\
& +(1-q)\int \sum_i 1_{\{S''_{W,i}\in(u-v,u-v+h)\}}f(v)\,dv)] \qquad (5.4.10) \\
= \; & E[S_W(0,h)(q\int S''_B(u-v,u-v+h)f(v)\,dv \\
& +(1-q)\int S''_W(u-v,u-v+h)f(v)\,dv)] \qquad (5.4.11)
\end{aligned}
$$

where (5.4.9) is obtained by using the fact that the V_i are i.i.d. with density $f(\cdot)$. As in the derivation of the expression for p_{R_W}, by application of the Tonelli-Fubini theorem, the integral and sum are reversed to get (5.4.10).

Since S_W, S_B, S''_W, and S''_B are counting measures (5.4.11) may be expressed as

$$
\begin{aligned}
& qE[\int_0^h S_W(dx_1)\int\int_u^{u+h} S''_B(d(y_1-v))f(v)\,dv] \\
& +(1-q)E[\int_0^h S_W(dx_1)\int\int_u^{u+h} S''_W(d(y_1-v))f(v)\,dv].
\end{aligned}
$$

By noting that the integrand is non-negative the order of integration and expectation may be reversed yielding

$$E[S_W(0,h)R_W(u,u+h)] = q\int_u^{u+h}\int_0^h\int E[S_W(dx_1)S''_B(d(y_1-v))]f(v)\,dv$$
$$+(1-q)\int_u^{u+h}\int_0^h\int E[S_W(dx_1)S''_W(d(y_1-v))]f(v)\,dv. \qquad (5.4.12)$$

Because of the thinning method, the two expectations in (5.4.12) can be expressed as

$$E[S_W(dx_1)S''_B(d(y_1-v))]$$
$$= E[S_W(dx_1)S'_B(d(y_1-v))1_{\{U_{y_1-v}>p\}}1_{\{S'_B(y_1-v,y_1-v+d)=0\}}], \qquad (5.4.13)$$

and

$$E[S_W(dx_1)S''_W(d(y_1-v))]$$
$$= E[S_W(dx_1)S'_W(d(y_1-v))1_{\{U_{y_1-v}>p\}}1_{\{S'_W(y_1-v,y_1-v+d)=0\}}]. \qquad (5.4.14)$$

Using the independence of U_{y_1-v} and the fact that S'_B and S_W are independent Poisson processes with rates $p_{S_B}+p_{N_B}$ and p_{S_W}, respectively, (5.4.13) can be re-written as

$$E[S_W(dx_1)S''_B(d(y_1-v))] = (1-p)e^{-d(p_{S_B}+p_{N_B})}p_{S_W}(p_{S_B}+p_{N_B})\,dx_1\,dy_1. \qquad (5.4.15)$$

To simplify (5.4.14), we note that if $0 < x_1-(y_1-v) < d$, then $S_W(y_1-v)$ is deleted, because there is then a point at least d time units after $S_W(y_1-v)$. Thus, we get the expression

$$E[S_W(dx_1)S''_W(d(y_1-v))]$$
$$= (1-p)e^{-d(p_{S_W}+p_{N_W})}(C_0(x_1,y_1,v)+C_1(x_1,y_1,v))\,dx_1\,dy_1, \qquad (5.4.16)$$

where

$$C_0(x_1,y_1,v) = p_{S_W}(p_{S_W}+p_{N_W})1_{\{x_1\notin(y_1-v,y_1-v+d)\}},$$
$$C_1(x_1,y_1,v) = p_{S_W}\delta(y_1-v-x_1).$$

The terms C_0 and C_1 account for the events that the condition $0 < x_1-(y_1-v) < d$ fails and holds, respectively. If the condition holds, then (5.4.16) is nonzero only when $x_1 = y_1-v$, which is identified by the Dirac delta in C_1.

Substitution of (5.4.15) and (5.4.16) into (5.4.12) gives

$$\begin{aligned}
&E[S_W(0,h)R_W(u,u+h)] \\
&= q\int_0^h\int_u^{u+h}\int (1-p)e^{-d(p_{S_B}+p_{S_B})}p_{S_W}(p_{S_B}+p_{S_B})f(v)\,dv\,dy_1\,dx_1 \\
&\quad +(1-q)\int_0^h\int_u^{u+h}\int (1-p)e^{-d(p_{S_W}+p_{N_W})}\Big[C_0(x_1,y_1,v)f(v)\,dv \\
&\qquad + C_1(x_1,y_1,v)f(v)\,dv\Big]\,dy_1\,dx_1 \qquad (5.4.17)
\end{aligned}$$

Using the property of the Dirac delta with $x = v$ and $x_0 = y_1 - x_1$, and noting that since $f(v)$ is a density function $\int f(v)\,dv = 1$, (5.4.17) simplifies to

$$\begin{aligned}
&E[S_W(0,h)R_W(u,u+h)] = qh^2(1-p)e^{-d(p_{S_B}+p_{S_B})}(p_{S_B}+p_{N_B})p_{S_W} \\
&\quad +(1-q)(1-p)e^{-d(p_{S_W}+p_{N_W})}p_{S_W}(p_{S_W}+p_{N_W}) \\
&\quad \times\int_0^h\int_u^{u+h}\left[\int 1_{\{v\notin(y_1-x_1,y_1-x_1+d)\}}f(v)\,dv + \frac{f(y_1-x_1)}{p_{S_W}+p_{N_W}}\right]dy_1\,dx_1. \qquad (5.4.18)
\end{aligned}$$

Finally, substituting (5.4.18), after two applications of the mean value theorem for integrals to the double integral involving $f(y_1 - x_1)$, we get an expression for $p_{S_W R_W}(u)$, as in (5.3.3):

$$\begin{aligned}
p_{S_W R_W}(u) &= \lim_{h\to 0} E[S_W(0,h)R_W(u,u+h)] \\
&= (1-p)p_{S_W}[q(p_{S_B}+p_{N_B})e^{-d(p_{S_B}+p_{N_B})} \\
&\quad +(1-q)e^{-d(p_{S_W}+p_{N_W})}\{(p_{S_W}+p_{N_W})P(V\notin(u,u+d))+f(u)\}].
\end{aligned}$$

5.4.3 Response-Response Second Order Intensities

We now look at $p_{R_W R_B}(u)$. Arguments similar to those in the derivation of the second-order stimulus-response second-order intensities lead to

$$
\begin{aligned}
&E[R_W(x,x+h)R_B(y,y+h)] \\
&= E[\{q\int\int_x^{x+h} S''_B(d(x_1-v_1))f(v_1)\,dv_1 + (1-q)\int\int_x^{x+h} S''_W(d(x_1-v_1))f(v_1)\,dv_1\} \\
&\quad \times\{q\int\int_y^{y+h} S''_W(d(y_1-v_2))f(v_2)\,dv_2 + (1-q)\int\int_y^{y+h} S''_B(d(y_1-v_2))f(v_2)\,dv_2\}] \\
&= (1-p)^2\Big[q(1-q)\int\int\int_x^{x+h}\int_{y:y_1\neq x_1}^{y+h} E[S'_B(d(x_1-v_1))S'_B(d(y_1-v_2)) \\
&\quad \times 1_{\{S'_B(x_1-v_1,x_1-v_1+d)=0\}}1_{\{S'_B(y_1-v_2,y_1-v_2+d)=0\}}]f(v_1)f(v_2)\,dx_1\,dy_1\,dv_1\,dv_2 \\
&\quad +p_{R_B}|(x,x+h)\bigcap(y,y+h)|_L \\
&\quad +q(1-q)\int\int\int_x^{x+h}\int_{y:y_1\neq x_1}^{y+h} E[S'_W(d(x_1-v_1))S'_W(d(y_1-v_2))1_{\{S'_W(x_1-v_1,x_1-v_1+d)=0\}} \\
&\quad \times 1_{\{S'_W(y_1-v_2,y_1-v_2+d)=0\}}]f(v_1)f(v_2)\,dx_1\,dy_1\,dv_1\,dv_2 + p_{R_W}|(x,x+h)\bigcap(y,y+h)|_L \\
&\quad +q^2\int\int\int_x^{x+h}\int_y^{y+h} E[S'_W(d(x_1-v_1))S'_B(d(y_1-v_2))1_{\{S'_W(x_1-v_1,x_1-v_1+d)=0\}} \\
&\quad \times 1_{\{S'_B(y_1-v_2,y_1-v_2+d)=0\}}]f(v_1)f(v_2)\,dx_1\,dy_1\,dv_1\,dv_2 \\
&\quad +(1-q)^2\int\int\int_x^{x+h}\int_y^{y+h} E[S'_B(d(x_1-v_1))S'_W(d(y_1-v_2))1_{\{S'_B(x_1-v_1,x_1-v_1+d)=0\}} \\
&\quad \times 1_{\{S'_W(y_1-v_2,y_1-v_2+d)=0\}}]f(v_1)f(v_2)\,dx_1\,dy_1\,dv_1\,dv_2\Big]
\end{aligned}
$$

Making use of the fact that S'_B and S'_W are independent Poisson processes, after division by h^2 and letting $h \to 0$, we obtain an expression for $p_{R_W R_B}(u)$, where $u = y - x$.

$$
\begin{aligned}
p_{R_W R_B}(u) &= (1-p)^2[P(|V_2 - V_1 - u| > d)\{q(1-q)(p_{S_B}+p_{N_B})^2 e^{-2d(p_{S_B}+p_{N_B})} \\
&+ q(1-q)(p_{S_W}+p_{N_W})^2 e^{-2d(p_{S_W}+p_{N_W})}\} \\
&+ q^2(p_{S_B}+p_{N_B})(p_{S_W}+p_{N_W})e^{-d(p_{S_B}+p_{N_B}+p_{S_W}+p_{N_W})} \\
&+ (1-q)^2(p_{S_B}+p_{N_B})(p_{S_W}+p_{N_W})e^{-d(p_{S_B}+p_{N_B}+p_{S_W}+p_{S_B})}] \qquad (5.4.19)
\end{aligned}
$$

The intensity $p_{R_B R_W}(u)$ is identical to $p_{R_W R_B}(u)$ by definition. In a similar manner

we obtain expressions (5.3.7) and (5.3.8) for $p_{R_W R_W}(u)$ and $p_{R_B R_B}(u)$, respectively.

5.5 Nonparametric Intensity Estimation

The approach to nonparametric intensity estimation is similar to that used for the go-no go RT experiment. First-order response intensities and stimulus-response intensities are estimated as in Section 4.5. Estimates of the response-response intensities are given by:

$$\hat{p}_{R_s R_r}(u) = \begin{cases} \frac{1}{Th}\sum_i \sum_j K(\frac{u-(R_{s,j}-R_{s,i})}{h}), & s \neq r \\ \frac{1}{Th}\sum_i \sum_{j\neq i} K(\frac{u-(R_{r,j}-R_{r,i})}{h}), & s = r \end{cases}.$$

5.5.1 Bandwidth Selection

We find optimal bandwidths for the intensities by minimizing the asymptotic mean integrated squared error (AMISE) with respect to h, the bandwidth. A general expression of the AMISE optimal bandwidth for a second-order intensity is derived in Section 2.6.1, and in the case of the biweight kernel we have

$$h_{AMISE}(p_{SR}) = \left[\frac{35\int p_{SR}(u)\,du}{T\int(\frac{\partial^2}{\partial^2 u}p_{SR}(u))^2\,du}\right]^{1/5}. \tag{5.5.1}$$

Using our fitted model, we can obtain estimates of $h_{AMISE}(p_{SR})$. Expressions for the stimulus-response second order intensities are given in (5.3.3),(5.3.4), (5.3.5), and (5.3.6) from which it can be seen that the terms involving u in the expressions with the same stimulus type are identical. For example, $p_{S_W R_W}(u)$ and $p_{S_W R_B}(u)$ depend on u only through

$$w(u) = P(V \notin (u, u+d)) + \frac{f(u)}{p_{S_W} + p_{N_W}},$$

which, under the assumption of a normal delay distribution, has second derivative

$$w''(u) = \frac{u+d-\mu}{\sigma^2}f(u+d) + \left[\frac{1}{(p_{S_W}+p_{N_W})\sigma^2}\left(\frac{(u-\mu)^2}{\sigma^2}-1\right) - \frac{(u-\mu)}{\sigma^2}\right]f(u),$$

where $f(\cdot)$ is the Normal density with mean μ and variance σ^2. Thus

$$\frac{\partial^2}{\partial^2 u}p_{S_W R_W}(u) = (1-p)(1-q)p_{S_W}(p_{S_W}+p_{N_W})e^{-d(p_{S_W}+p_{N_W})}w''(u),$$

$$\frac{\partial^2}{\partial^2 u} p_{S_B R_B}(u) = (1-p)(1-q) p_{S_B}(p_{S_B} + p_{N_B}) e^{-d(p_{S_B}+p_{N_B})} w''(u),$$

and

$$\frac{\partial^2}{\partial^2 u} p_{S_W R_B}(u) = (1-p) q p_{S_W}(p_{S_W} + p_{N_W}) e^{-d(p_{S_W}+p_{N_W})} w''(u).$$

Analogous expressions can be found for $p_{S_B R_B}(u)$ and $p_{S_B R_W}(u)$.

Expressions for the response-response second order intensities are given in (5.3.7), (5.3.8), and (5.3.9). It can be seen that each intensity depends on u by the same term:

$$r(u) = P(V_2 - V_1 - u| > d),$$

which has second derivative

$$r''(u) = \frac{1}{2\sigma^2}\Big[(u+d)g(u+d) - (u-d)g(u-d)\Big],$$

where $g(\cdot)$ denotes the Normal density with mean 0 and variance $2\sigma^2$. Thus,

$$\begin{aligned}\frac{\partial^2}{\partial^2 u} p_{R_W R_W}(u) = (1-p)^2 [q^2 (p_{S_B} + p_{N_B})^2 e^{-2d(p_{S_B}+p_{N_B})} \\ + (1-q)^2 (p_{S_W} + p_{N_W})^2 e^{-2d(p_{S_W}+p_{N_W})}] r''(u),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial^2 u} p_{R_B R_B}(u) = (1-p)^2 [q^2 (p_{S_W} + p_{N_W})^2 e^{-2d(p_{S_W}+p_{N_W})} \\ + (1-q)^2 (p_{S_B} + p_{N_B})^2 e^{-2d(p_{S_B}+p_{N_B})}] r''(u),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial^2 u} p_{R_W R_B}(u) = (1-p)^2 [q(1-q)(p_{S_B} + p_{N_B})^2 e^{-2d(p_{S_B}+p_{N_B})} \\ + q(1-q)^2 (p_{S_W} + p_{N_W})^2 e^{-2d(p_{S_W}+p_{N_W})}] r''(u).\end{aligned}$$

Finally, for a given second-order intensity, substitution of the intensity expression and corresponding second derivative into (5.5.1) yields the exact form of an optimal bandwidth.

5.6 Parameter Estimation

Assumptions 1 through 5 allow us to use a quick approximation method that is a modification of the one used by Braun et al. (2003) to estimate μ, σ, and the total noise process rate $p_{N_B} + p_{N_W}$, which we denote by $p_{N_{B,W}}$.

For estimation, we also assume that the black and white noise process rates are proportional to their respective flash rates, so that

$$p_{N_B} = p_{N_{B,W}} \frac{p_{S_B}}{p_{S_B} + p_{S_W}},$$

and

$$p_{N_W} = p_{N_{B,W}} \frac{p_{S_W}}{p_{S_B} + p_{S_W}}.$$

It is reasonable to assume that when flashes are presented at the same rate, the rates of the noise processes will also be similar. Likewise, when one flash rate is higher than the other, it is reasonable to assume that the noise process corresponding to the faster flash process has a faster rate than the other noise process; there is likely more noise associated with the flash process with a higher rate.

In Braun et al. (2003) an iterative algorithm is used to obtain $\hat{\mu}$, $\hat{\sigma}$, and the total noise rate for simple RT data. Assuming independent and identically distributed reaction times from both the black and white flashes allows us to consider only the times of the flashes and responses, and not the types of flashes and responses. Since the locations of the peaks of the intensities $p_{S_W R_W}(u)$ and $p_{S_B R_B}(u)$ correspond to the modes of the white and black reaction times, respectively, and we are assuming identical symmetric distributions, the locations of the peaks for the corresponding nonparametric estimates should be near the value of the mean reaction time. Therefore, as a modified algorithm, we fix the mean of the two peak locations, μ_0, as an initial value for the estimation of μ. The estimates of σ and the total noise rate $p_{N_B} + p_{N_W}$ are as in their algorithm. The modified algorithm is given as Algorithm 5.6.1. The algorithm is dependent on the assumption that the reaction times are normally distributed, but it can be adapted to handle other distributions, if necessary.

Algorithm 5.6.1. *For RT data (S, R) the parameters μ, σ, p_{N_B}, and p_{N_W} can be estimated as follows:*

1. *Find the location of the peak in nonparametric intensity estimates of $p_{S_W R_W}(u)$ and $p_{S_B R_B}(u)$ and set μ_0 as the mean of the two locations.*

2. *Match each response time R_j with a flash time preceding it, $\hat{S}_j''$, such that $R_j - \hat{S}_j''$ is closest to μ_0, to obtain estimates of the reaction times: $\hat{V}_j = R_j - \hat{S}_j''$.*

3. *Estimate μ and σ by the mean and standard deviation of the $\hat{V}_j$.*

4. *Let $T = R_{N_R}$, where N_R is the number of responses. Estimate the total noise rate $p_{N_{B,W}}$, by:*

$$\hat{p}_{N_{B,W}} = \frac{|\{\hat{S}_j'' : \hat{S}_j'' = \hat{S}_k''\}|}{T},$$

where $|\cdot|$ denotes set cardinality.

5. *Estimate individual noise rates by:*

$$\hat{p}_{N_B} = \hat{p}_{N_{B,W}} \frac{p_{S_B}}{p_{S_B} + p_{S_W}},$$

and

$$\hat{p}_{N_W} = \hat{p}_{N_{B,W}} \frac{p_{S_W}}{p_{S_B} + p_{S_W}}.$$

Examination of plots of the intensities $p_{S_W R_B}(u)$ (and $p_{S_B R_W}(u)$) and $p_{R_W R_W}(u)$ (and $p_{R_B R_B}(u)$) indicate that these functions may be useful in obtaining estimates of q and d, respectively. Furthermore, for long-run experiments (large T values) the bias of each of the nonparametric intensity estimates is near zero, and one expects the difference between the nonparametric and parametric estimates to be minimal. Let $\hat{p}_{S_W R_B}(\hat{\mu})$ and $\hat{p}_{S_W R_B,h}(\hat{\mu})$ denote the parametric and the nonparametric (with bandwidth h) estimates, respectively. We obtain estimates of d, q and p by minimizing the differences between parametric (based on model expression with parameter estimates) and corresponding nonparametric estimates, evaluated at a certain point. For each of the four cases resulting from whether or not the stimulus rates are equal, and the absence or presence of nonlinear inhibition, we develop an estimation method. The stimulus rates are equal for the choice RT experiment data, but for completeness, we also provide estimation methods for the more general cases involving different stimulus rates.

Recall the expressions for p_{R_W} and p_{R_B}, given in equations (5.3.1) and (5.3.2), respectively. Manipulating each of these equations, with p_{R_W} and p_{R_B} replaced by

their nonparametric estimates, yields an expression for $\hat{p}$ that involves $\hat{q}$ and $\hat{d}$. We denote these two expressions for $\hat{p}$ by $\hat{p}_W$ and $\hat{p}_B$, which are given as follows, with p_{N_B} and p_{N_W} replaced by their estimates obtained from Algorithm 5.6.1:

$$\hat{p}_W = 1 - \frac{\hat{p}_{R_W}}{q(p_{S_B} + \hat{p}_{N_B})e^{-\hat{d}(p_{S_B}+\hat{p}_{N_B})} + (1-\hat{q})(p_{S_W} + \hat{p}_{N_W})e^{-\hat{d}(p_{S_W}+\hat{p}_{N_W})}} \tag{5.6.1}$$

and

$$\hat{p}_B = 1 - \frac{\hat{p}_{R_B}}{\hat{q}(p_{S_W} + \hat{p}_{N_W})e^{-\hat{d}(p_{S_W}+\hat{p}_{N_W})} + (1-\hat{q})(p_{S_B} + \hat{p}_{N_B})e^{-\hat{d}(p_{S_B}+\hat{p}_{N_B})}}. \tag{5.6.2}$$

We estimate p by the simple average $(\hat{p}_W + \hat{p}_B)/2$, which we denote $\hat{p}$, and our estimation problem is reduced to finding $\hat{q}$ and $\hat{d}$. When the stimulus rates are equal and/or $d = 0$, the expression for $\hat{p}$ is reduced to a simpler form, and a simpler estimation procedure can be used. The different cases and corresponding estimation methods are given in the following subsections.

5.6.1 Equal Stimulus Rates and Absence of Nonlinear Inhibition

First, we consider estimation for the simplest case of equal stimulus rates and no nonlinear inhibition, so that $p_{S_B} = p_{S_W} := p_S$ and $d = 0$. By assumption, since the stimulus rates are equal, we have $p_{N_B} = p_{N_W} := p_N$. In this case, our estimate of p becomes

$$\hat{p} = 1 - \frac{1}{2(p_S + \hat{p}_N)}(\hat{p}_{R_W} + \hat{p}_{R_B}), \tag{5.6.3}$$

which can be evaluated without knowledge of q because of the equal stimulus rates. Since $0 \leq p \leq 1$, in the event that $\hat{p} < 0$, we set $\hat{p} = 0$.

In this case, for a fixed u the expression for $p_{S_W R_B}(u)$ (see (5.3.4)) is a linear function of q:

$$p_{S_W R_B}(u) = (1-p)p_S[p_S + p_N + f(u)q]. \tag{5.6.4}$$

The estimation is done at $u = \hat{\mu}$, which is the location of the peak of $p_{S_W R_B}(u)$. Insertion of the nonparametric estimator $\hat{p}_{S_W R_B,h}(\mu)$ into (5.6.4) yields

$$\hat{q} = \left(\frac{\hat{p}_{S_W R_B,h}(\hat{\mu})}{(1-\hat{p})p_S(p_S + \hat{p}_N)} - 1\right)\frac{(p_S + \hat{p}_N)}{\phi_{\hat{\mu},\hat{\sigma}}(\hat{\mu})}, \tag{5.6.5}$$

where $\hat{p}$ is replaced by (5.6.3), and we assume a $N(\mu, \sigma)$ density, which is estimated using the estimates $\hat{\mu}, \hat{\sigma}$ obtained from Algorithm 5.6.1.

Since the estimation of q is dependent on $\hat{p}_{S_W R_B, h}(\hat{\mu})$, the accuracy of the q estimate will depend on the choice of bandwidth. As shown in the previous section, the AMISE optimal bandwidth depends on the values of the model parameters. Therefore, we first find an initial bandwidth h using the already obtained estimates $\hat{\mu}$ and $\hat{\sigma}$, along with $d = 0$, $q = q_0$ and $p = \hat{p}$. Since the error probability is likely relatively small, we use $q_0 = .05$, as an initial value. This h is then used in (5.6.5) to obtain the estimate $\hat{q}$, which is then used to update the bandwidth and evaluate (5.6.5) once again, yielding our final estimate $\hat{q}$, and finally $\hat{p}$.

5.6.2 Equal Stimulus Rates and Nonlinear Inhibition

Now we allow $d > 0$, so that there is nonlinear inhibition, and maintain the restriction of equal stimulus rates. In this case, our expression for $\hat{p}$ involves d, and is given by

$$\hat{p} = 1 - \frac{\hat{p}_{R_W} + \hat{p}_{R_B}}{2(p_S + \hat{p}_N)e^{-\hat{d}(p_S + \hat{p}_N)}}. \tag{5.6.6}$$

Based on the expression for $p_{S_W R_B}(u)$, knowledge of p can be used to obtain $\hat{q}$ as follows:

$$\hat{q} = \left(1 - \frac{\hat{p}_{S_W R_B, h}(\hat{\mu}) e^{\hat{d}(p_S + \hat{p}_N)}}{(1 - \hat{p}) p_S (p_S + \hat{p}_N)}\right) \frac{1}{1 - P(V \notin (u, u + d)) - \phi_{\hat{\mu}, \hat{\sigma}}(\hat{\mu})/(p_S + \hat{p}_N)}. \tag{5.6.7}$$

Consequently, upon substitution of (5.6.6) into (5.6.7) we obtain:

$$\hat{q} = \left(1 - \frac{2\hat{p}_{S_W R_B, h}(\hat{\mu})}{p_S(\hat{p}_{R_W} + \hat{p}_{R_B})}\right) \frac{1}{1 - P(V \notin (u, u + d)) - \phi_{\hat{\mu}, \hat{\sigma}}(\hat{\mu})/(p_S + \hat{p}_N)}. \tag{5.6.8}$$

Thus, $\hat{p}$ and $\hat{q}$ can be found upon obtaining an estimate of d. In the event that $\hat{p}$ or $\hat{q}$ is negative, the corresponding estimate is set to zero.

As discussed earlier either $p_{R_W R_W}(u)$ or $p_{R_B R_B}(u)$ give an indication of the magnitude of d. In the estimation of d we use $p_{R_W R_W}(u)$ (or $p_{R_B R_B}(u)$), expressing p and q as in (5.6.6) and (5.6.8), respectively. A closed-form expression for d cannot be found, so we minimize with respect to d the difference between the nonparametric and parametric estimates of $p_{R_W R_W}(0)$ (or $p_{R_B R_B}(0)$); the location of the function's

trough is at $u = 0$. This is accomplished by using the `optimize` function in the statistical package R (R Development Core Team (2006)), which uses a combination of golden section search and successive parabolic interpolation. Again, because of the dependence on the nonparametric estimator, the estimation is done in two steps in order to update the bandwidth and improve our estimates. However, if $\hat{q} = 0$ after the first step, the intensity $p_{R_W R_W}(u)$ (or $p_{R_B R_B}(u)$) is constant according to our model, and there is no need to perform the second step.

5.6.3 Different Stimulus Rates and Absence of Nonlinear Inhibition

Next, consider the case where the stimulus rates are different, and the absence of nonlinear inhibition, so that $d = 0$. In this case our estimate of p involves q and is given by:

$$\hat{p} = 1 - \frac{1}{2}\left[\frac{\hat{p}_{R_W}}{\hat{q}(p_{S_B} + \hat{p}_{N_B}) + (1-\hat{q})(p_{S_W} + \hat{p}_{N_W})} + \frac{\hat{p}_{R_B}}{\hat{q}(p_{S_W} + \hat{p}_{N_W}) + (1-\hat{q})(p_{S_B} + \hat{p}_{N_B})}\right], \tag{5.6.9}$$

and $p_{S_W R_B}(u)$ has the form:

$$\begin{aligned} p_{S_W R_B}(u) = (1-p)p_{S_W}[p_{S_B} + p_{N_B} \\ + q\{(p_{S_W} + p_{N_W}) + f(u) - (p_{S_B} + p_{N_B})\}]. \end{aligned} \tag{5.6.10}$$

Substitution of (5.6.9) into (5.6.10) yields a long and complicated quadratic in q. The expressions for the two solutions can be found using Mathematica (Stephen Wolfram (2006)). The estimation is done at $u = \hat{\mu}$, which is the location of the peak of $p_{S_W R_B}(u)$, and as in the other cases the estimation is done in two steps to improve our estimates. Of the two solutions for q from the quadratic, one tends to be large and negative, while the other tends to be between 0 and 1. We set our estimate of q as the solution that falls in $[0, 1]$. In the case that neither solution is in $[0, 1]$, we set $\hat{q} = 0$; one solution is very small and negative in this case.

5.6.4 Different Stimulus Rates and Nonlinear Inhibition

For the most general case, the stimulus rates are unequal and $d > 0$, so that there is nonlinear inhibition. In this case knowledge of both d and q can be used to evaluate $\hat{p} = (\hat{p}_W + \hat{p}_B)/2$, and $p_{S_W R_B}(u)$ has the form:

$$p_{S_W R_B}(u) = (1-p)p_{S_W}[(p_{S_B} + p_{N_B})e^{-d(p_{S_B}+p_{N_B})} + qe^{-d(p_{S_W}+p_{N_W})}\{(p_{S_W} + p_{N_W})P(V \notin (u, u+d)) + f(u) - (p_{S_B} + p_{N_B})\}] \quad (5.6.11)$$

Substitution of the expression for $\hat{p}$ into (5.6.11) yields a quadratic equation of q, for which knowledge of d is required for evaluation. The evaluation is done at $u = \hat{\mu}$, which is the location of the peak of $p_{S_W R_B}(u)$. As in the previous case, the quadratic has a long and complicated form, and the expressions for the two solutions can be found using Mathematica (Stephen Wolfram (2006)). Note that the solutions both require knowledge of d. For a given d, we set our estimate of q as the solution that falls in $[0, 1]$. In the case that neither solution is in $[0, 1]$, we set $\hat{q} = 0$; one solution is very small and negative in this case.

In the estimation of d we use $p_{R_W R_W}(u)$ (or $p_{R_B R_B}(u)$), expressing p and q as functions of d. A closed-form expression for d cannot be found, so we minimize with respect to d the difference between the nonparametric and parametric estimates of $p_{R_W R_W}(0)$ (or $p_{R_B R_B}(0)$); the location of the function's trough is at $u = 0$. As in the case of equal stimulus rates, $\hat{d}$ can be found by using the `optimize` function in the statistical package R. Again, because of the dependence on the nonparametric estimator, the estimation is done in two steps in order to update the bandwidth and improve our estimates. If $\hat{q} = 0$ after the first step, the second step is not implemented.

5.7 Applications to Simulated Data

In this section we study the behaviour of our intensity function estimates for simulated data. Simulated data is generated from our parametric choice RT model using Algorithm 5.2.1. The peak locations of the nonparametric estimates of $p_{S_W R_W}(u)$ and $p_{S_B R_B}(u)$ tend to coincide; the difference between the two estimates is usually at

most 0.05, for simulated data with $\mu \in [.4, .5]$, $\sigma \in [.1, .15]$, $d \in [0, .2]$, $p \in [.08, .15]$, $q \in [0, .5]$, $p_N = 0$.

5.7.1 Equal Stimulus Rates and Absence of Nonlinear Inhibition

Here, we consider the simplest type of choice RT data, which has $p_{S_B} = p_{S_W}$, and $d = 0$. For such data, we use the estimation method of Section 5.6.1, and we provide an example of the estimation for simulated data as follows.

Figure 5.4 displays the parametric and nonparametric stimulus-response intensity estimates for one set of simulated choice RT data with parameters $\mu = .4, \sigma = .1, d = 0, p = .15, p_{S_B} = .5, p_{S_W} = .5, p_N = 0$, and $q = .1$. The estimates obtained are $\hat{\mu} = .399, \hat{\sigma} = .091, \hat{p}_N = .027, \hat{p} = .225$, and $\hat{q} = .080$. The final bandwidth used in the estimation is $h = 0.22$.

The parametric and nonparametric intensity estimates correspond quite well. Similarly low peak heights are revealed in both plots of the intensity estimates involving flashes with responses of the opposite type. This agrees with the small q used in the simulation, and the correspondingly small $\hat{q}$. The variability of the estimates appears to be quite low. Table 5.1 provides the means and standard errors of estimates based on 500 simulations of 800 flashes with the same parameter settings as in the previous example. The estimates have small standard errors with the largest standard errors, .0281 and .0258, belonging to q and p, respectively.

Table 5.1: Standard errors of parameter estimates for simulated choice RT data with rate $p_S = .5$ and $d = 0$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.397	.0050
σ	.1	.091	.0026
q	.1	.077	.0281
p	.15	.202	.0258
p_N	0	.035	.0046

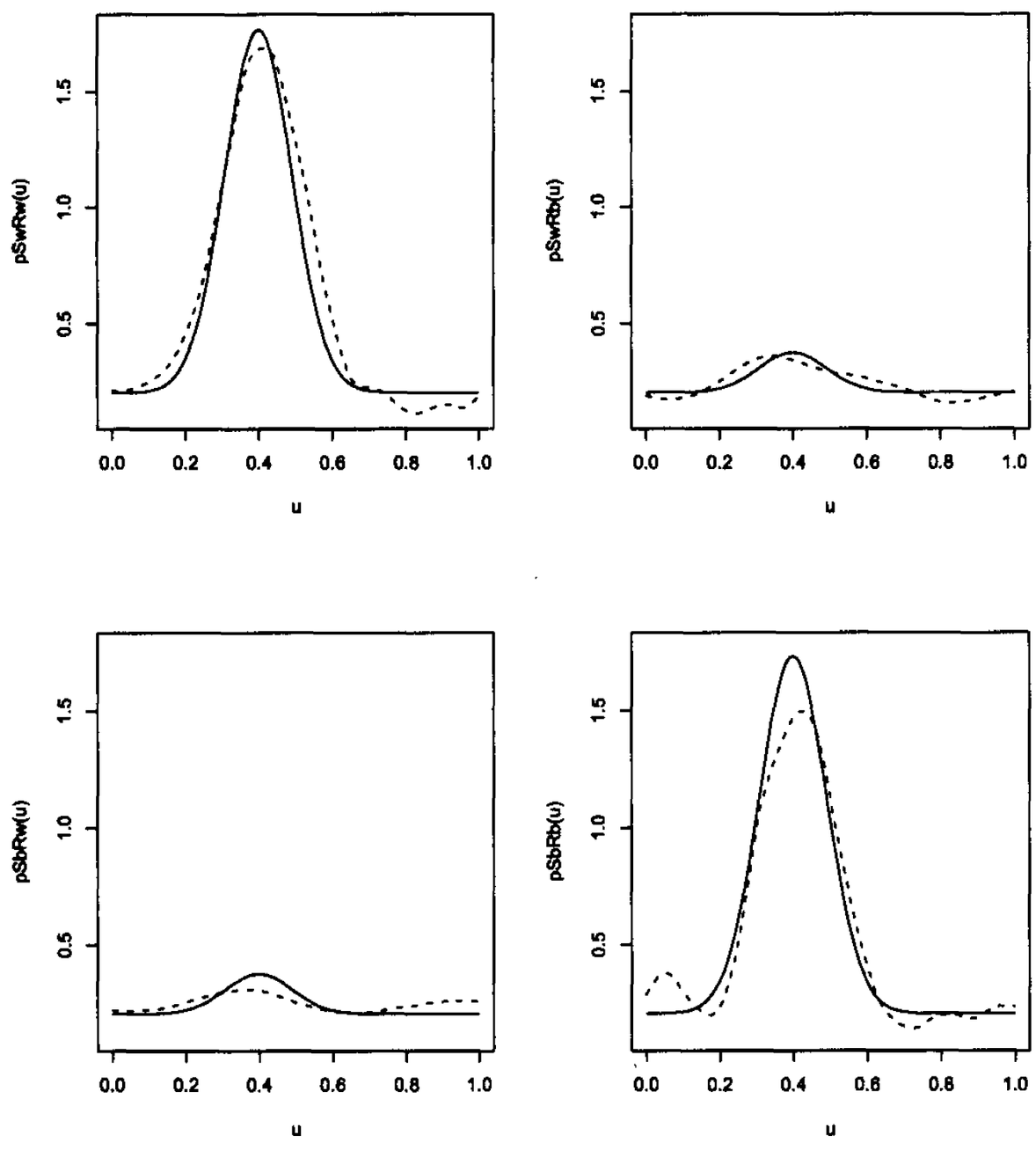


Figure 5.4: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated choice RT data with $p_{S_W} = .5$, $p_{S_B} = .5$, and $d = 0$.

5.7.2 Equal Stimulus Rates and Nonlinear Inhibition

For this type of choice RT data we still have the simplification of $p_{S_B} = p_{S_W}$, but now $d > 0$. For such data, we use the estimation method of Section 5.6.2. In that estimation method, either $p_{R_W R_W}(u)$ or $p_{R_B R_B}(u)$ can be used to obtain an estimate of d. Our estimation depends on the typical shape of the response-response intensities: a valley centered at 0, which widens and deepens as d increases. Occasionally, for some simulated data sets we find that one of the nonparametric estimates of $p_{R_W R_W}(u)$ and $p_{R_B R_B}(u)$ has a peak at 0 rather than a trough. This is an indication that the nonparametric estimates should be examined before parameter estimation in order to choose the more appropriate intensity function for the estimation of d.

An example where it is necessary to employ $p_{R_B R_B}(u)$ because of the behavior

of the nonparametric estimates follows. The true parameters for the data are $\mu = .5$, $\sigma = .133$,$d = .1$, $p = .11$, $p_{S_B} = p_{S_W} = .5$, $q = .025$, and $p_N = 0$. Using $p_{R_W R_W}(u)$,which has a peak at 0 for its nonparametric estimate, the $\hat{d}$ obtained is approximately 0. However, when $p_{R_B R_B}(u)$ is used we get $\hat{d} = .078$, which is close to the true parameter value $d = .1$. The other estimates are $\hat{\mu} = .484$, $\hat{\sigma} = .125$, $\hat{q} = 0$, $\hat{p} = .228$, and $\hat{p}_N = .034$.

Plots comparing the nonparametric and parametric estimates are displayed in Figures 5.5 and 5.6. The peak at lag 0 for the nonparametric estimate of $p_{R_W R_W}(u)$ indicates the necessity of using $p_{R_B R_B}(u)$ for the estimation of d. The stimulus-response parametric and nonparametric intensity estimates are very similar, as well as the cross response-response intensities $p_{R_W R_B}(u)$ and $p_{R_B R_W}(u)$. The $p_{R_B R_B}(u)$ estimates achieve the same depth, but the nonparametric estimate is much wider.

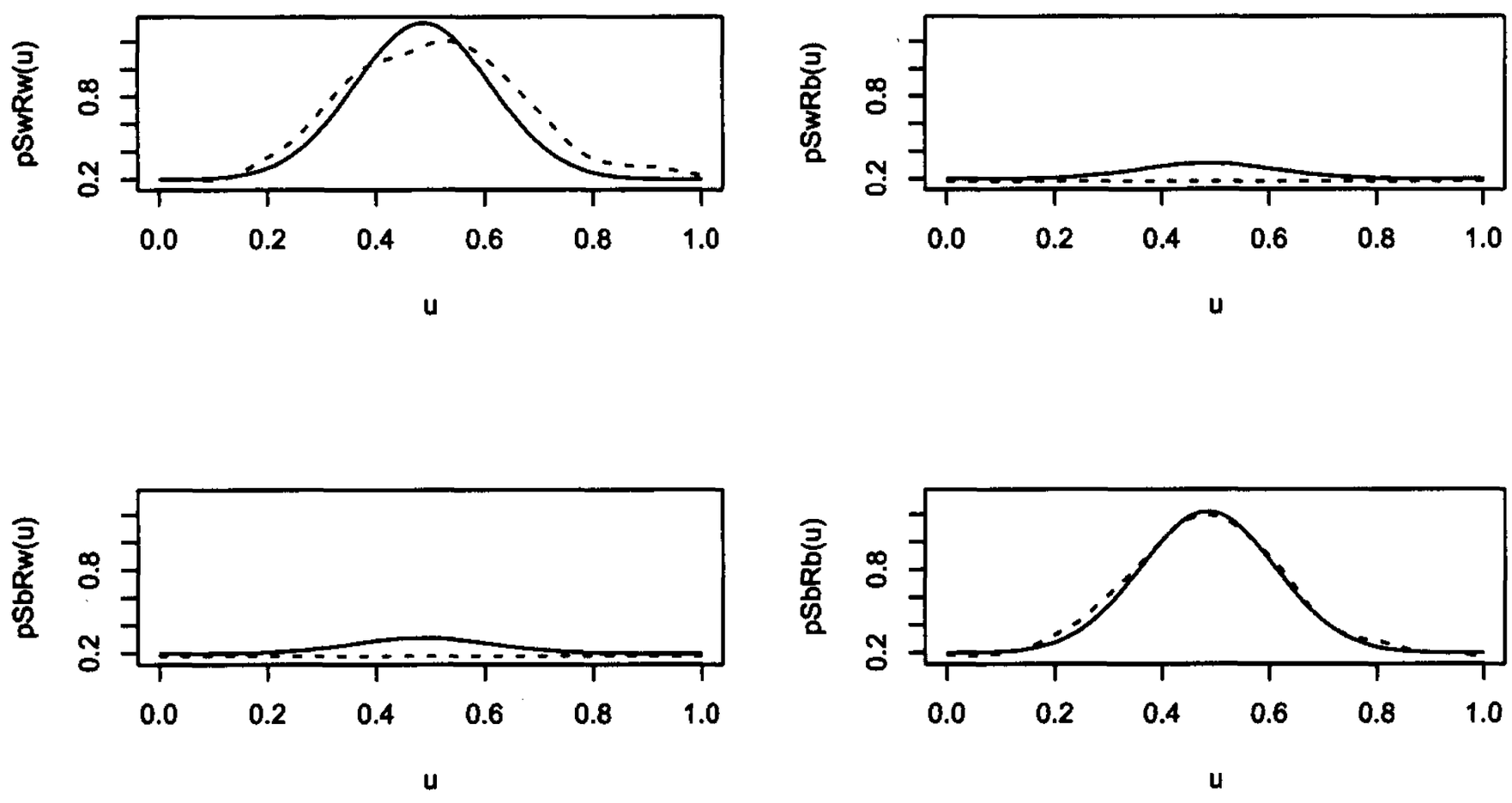


Figure 5.5: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .5$, and $d = .2$. Estimation employed $p_{R_B R_B}(u)$.

An example in which we employ $p_{R_W R_W}(u)$ follows. Figures 5.7 and 5.8 display the parametric and nonparametric stimulus-response and response-response intensity

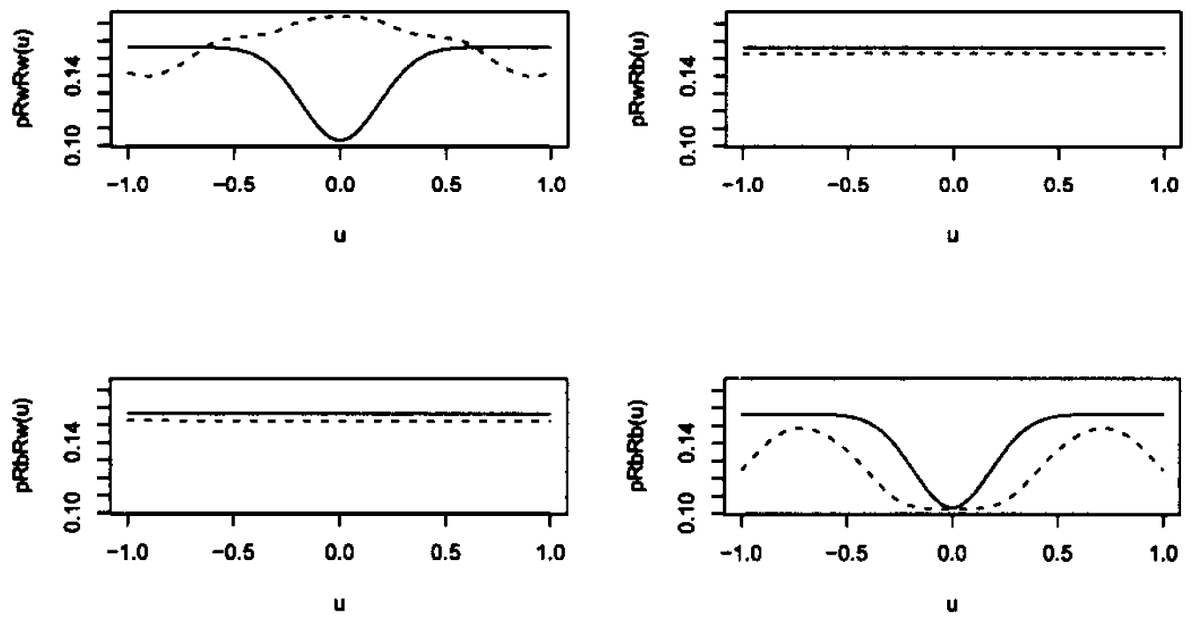


Figure 5.6: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .5$, and $d = .2$. Estimation employed $p_{R_B R_B}(u)$.

estimates for one set of simulated choice RT data with parameters $\mu = .4, \sigma = .1, d = .2, p = .15, p_{S_B} = .5, p_{S_W} = .5, p_N = 0$, and $q = .1$. The estimates obtained are $\hat{\mu} = .395, \hat{\sigma} = .087, \hat{p}_N = .018, \hat{p} = .167, d = .166$, and $\hat{q} = .075$. The two estimates of $p_{R_W R_W}(u)$ and the stimulus-response intensities coincide quite well.

Standard errors of the estimates for 500 simulations with parameters $\mu = .4, \sigma = .1, d = .2, p = .15, p_{S_B} = .5, p_{S_W} = .5, p_N = 0$, and $q = .1$ are given in Table 5.2. The largest standard error 0.0433 is for $\hat{d}$, which tends to be underestimated, on average.

Table 5.2: Standard errors of parameter estimates for simulated choice RT data with rate $p_S = .5$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.400	.0051
σ	.1	.092	.0029
d	.2	.141	.0433
q	.1	.072	.0285
p	.15	.199	.0316
p_N	0	.019	.0035

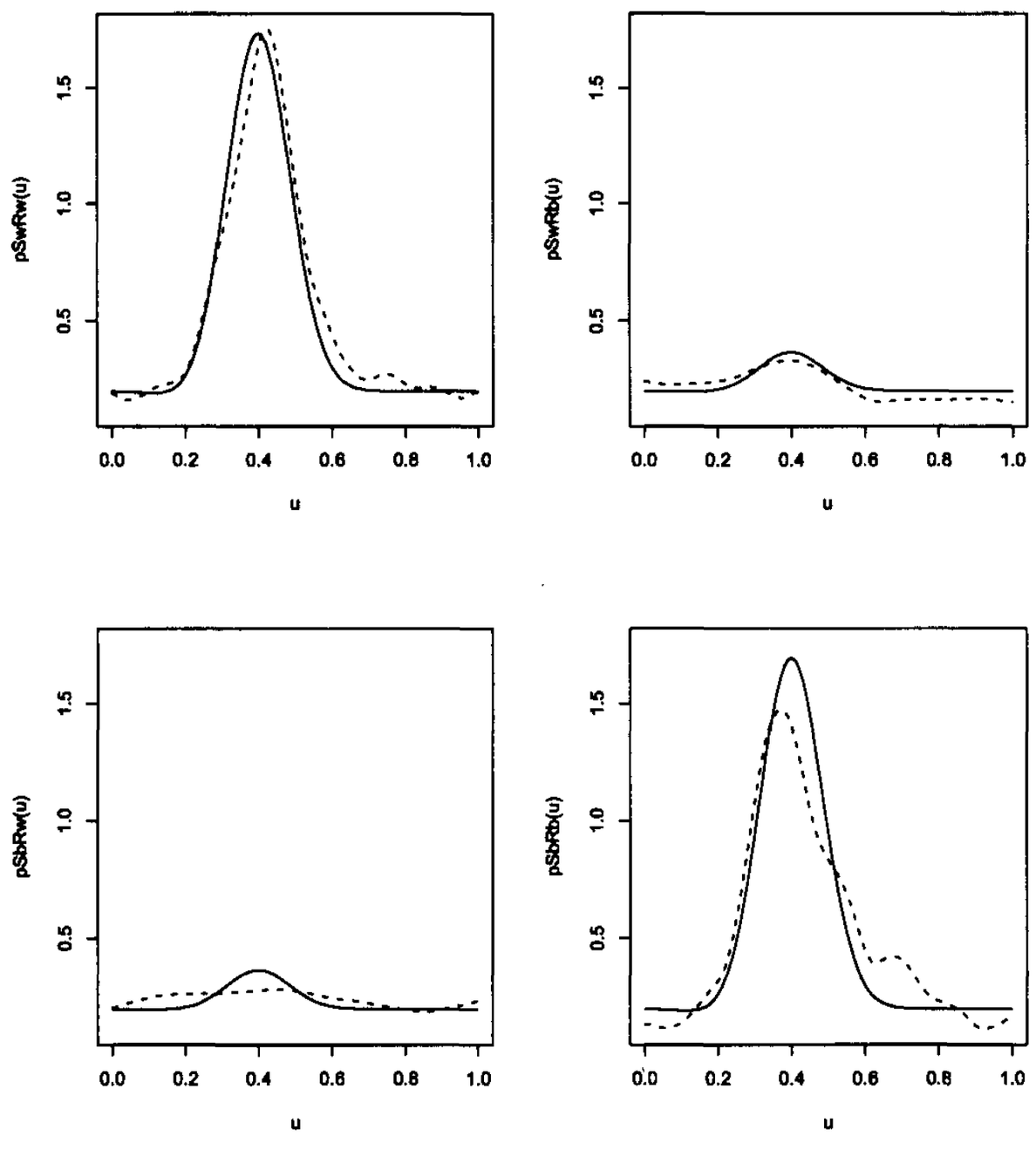


Figure 5.7: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .5$, and $d = .2$.

5.7.3 Different Stimulus Rates and Absence of Nonlinear Inhibition

We now consider the case where $p_{S_W} \neq p_{S_B}$, and $d = 0$. Our estimation method of Section 5.6.3 is implemented for such data, and we provide an example of the estimation for simulated data as follows.

Figure 5.9 displays the parametric and nonparametric stimulus-response intensity estimates for simulated choice RT data parameters $\mu = .4, \sigma = .1, d = 0, p = .15, p_{S_B} = .5, p_{S_W} = .7, p_N = 0$, and $q = .1$. The estimates obtained are $\hat{\mu} = .401, \hat{\sigma} = .085, \hat{p} = .206$, $\hat{q} = .114$, $p_{N_B} = .040$, and $p_{N_W} = .057$. Estimates of the stimulus-response intensities match very closely, except for a slightly lower peak in

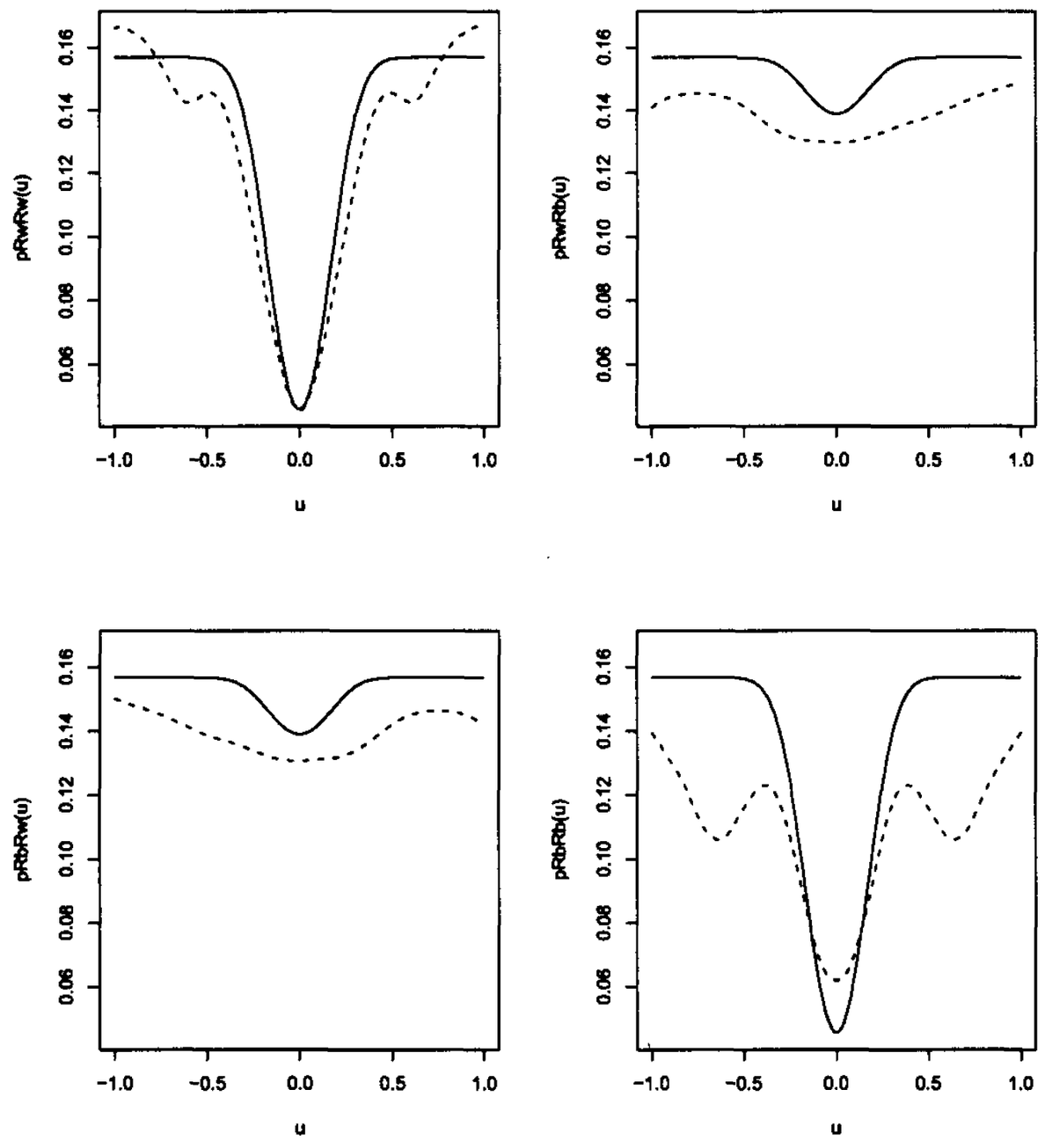


Figure 5.8: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .5$, $p_{S_B} = .5$, and $d = .2$.

the nonparametric estimate of $p_{S_B R_B}(u)$.

Standard errors of the estimates for 500 simulations with parameters $\mu = .4, \sigma = .1, d = 0, p = .15, p_{S_B} = .7, p_{S_W} = .5, p_N = 0$, and $q = .1$ are given in Table 5.3. The standard errors are quite small with $\hat{p}$ and $\hat{q}$ having the largest standard errors, .0267 and .0200, respectively.

5.7.4 Different Stimulus Rates and Nonlinear Inhibition

In the final case, we have $p_{S_W} \neq p_{S_B}$, and $d > 0$. The estimation method of Section 5.6.4 is implemented for such data, and we provide an example of the estimation for simulated data as follows.

Figures 5.10 and 5.11 display the parametric and nonparametric stimulus-response

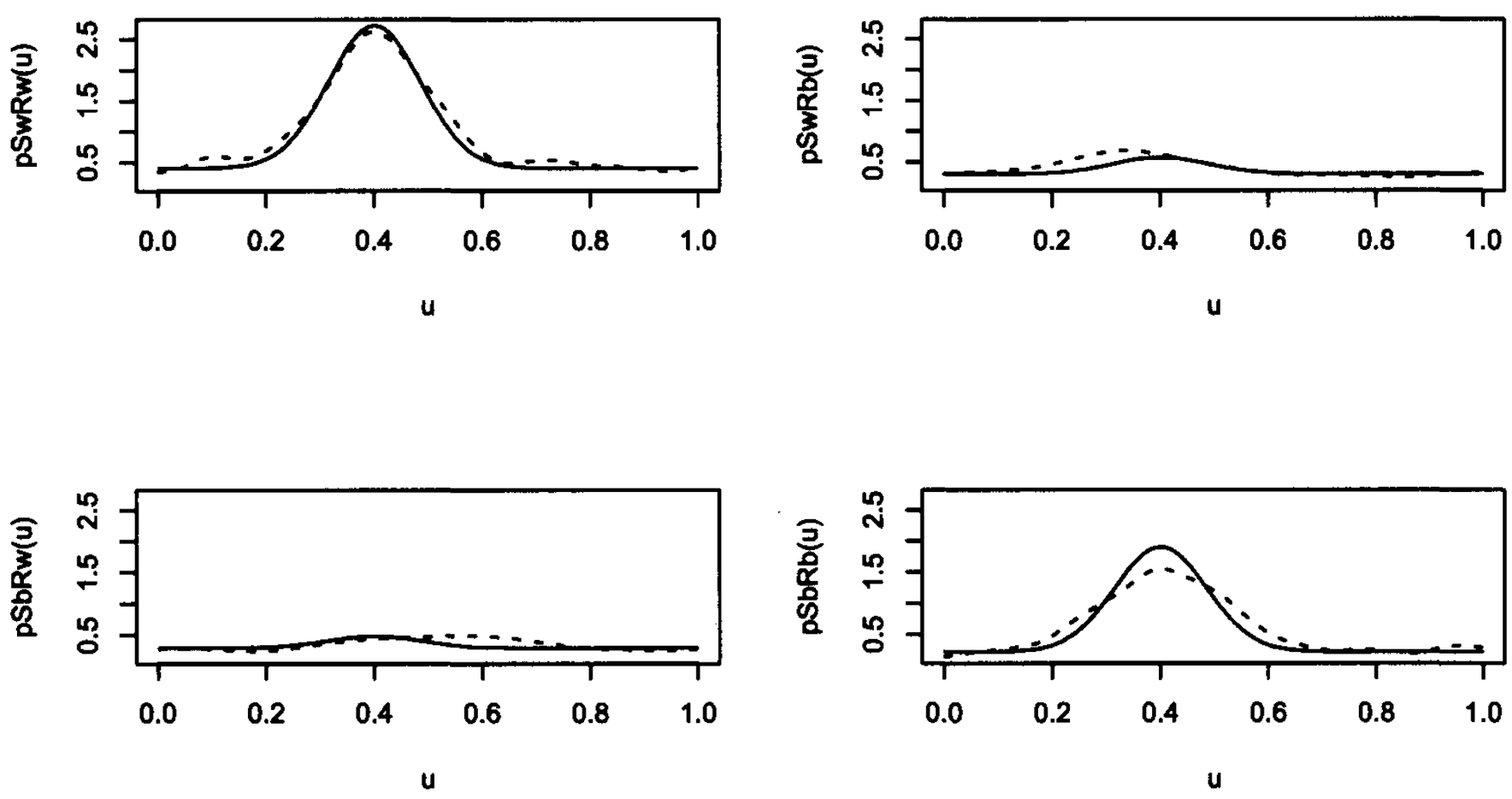


Figure 5.9: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .7$, $p_{S_B} = .5$, and $d = 0$.

and response-response intensity estimates for simulated choice RT data with parameters $\mu = .4, \sigma = .1, d = .2, p = .15, p_{S_B} = .5, p_{S_W} = .7, p_N = 0$, and $q = .1$. The estimates obtained are $\hat{\mu} = .406, \hat{\sigma} = .090, \hat{p} = .120, \hat{d} = .219$, $\hat{q} = .101$, $p_{N_B} = .023$, and $p_{N_W} = .033$. For each intensity, the parametric and nonparametric estimates are very similar, except in the case of $p_{R_B R_B}(u)$; the parametric estimate of $p_{R_B R_B}(u)$ has a deeper valley at lag 0 than the corresponding nonparametric estimate.

Table 5.4 displays the standard errors of the estimates for 500 simulations with parameters $\mu = .4, \sigma = .1, d = .2, p = .15, p_{S_B} = .7, p_{S_W} = .5, p_N = 0$, and $q = .1$. The estimates of d have the largest standard error, .0468, and tend to be underestimated.

5.8 An Application to Experimental Data

We fit our model to nine pooled runs (pooling based on 8 runs of 100 flashes each), each with equal stimulus rates. A description of how we pooled the data is provided in Section 3.3.6. We also consider an example of estimation for the individual unpooled

Table 5.3: Standard errors of parameter estimates for simulated choice RT data with $p_{S_W} = .7$, $p_{S_B} = .5$, and $d = 0$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.396	.0046
σ	.1	.089	.0026
q	.1	.073	.0200
p	.15	.210	.0267
p_{N_B}	0	.040	.0049
p_{N_W}	0	.056	.0070

Table 5.4: Standard errors of parameter estimates for simulated choice RT data with $p_{S_W} = .7$, $p_{S_B} = .5$, based on 500 simulations

parameter	true value	mean	standard error
μ	.4	.402	.0050
σ	.1	.090	.0029
d	.2	.152	.0468
q	.1	.074	.0273
p	.15	.199	.0380
p_{N_B}	0	.022	.0036
p_{N_W}	0	.030	.0050

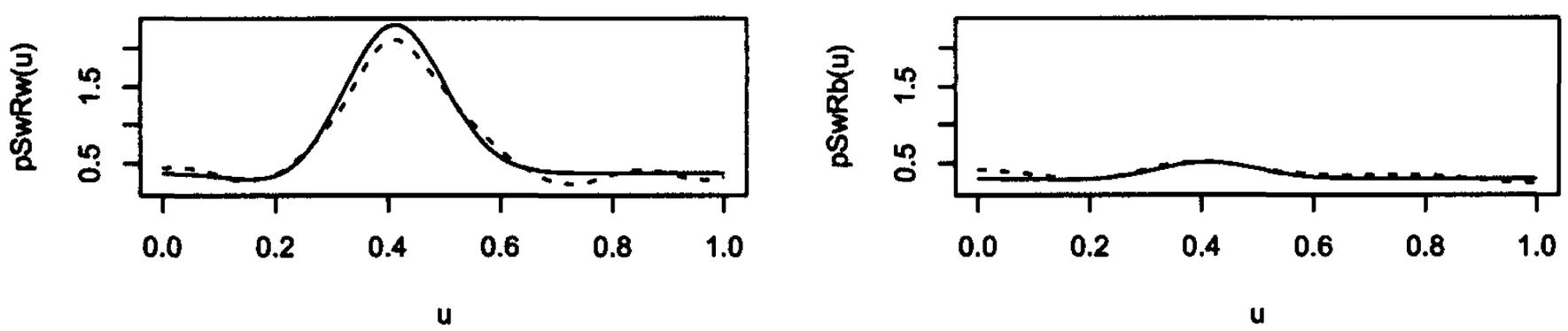


Figure 5.10: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .7$, $p_{S_B} = .5$, and $d = .2$.

data. The stimulus rates are 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 2.0, 4.0, and 8.0 flashes per second.

Occasionally there is a '2' instead of a '0' (black) or '1' (white) recorded in the response data. This means that both the black and white buttons were pressed almost simultaneously. We tried two different methods for dealing with such instances. First, we recorded the response times with a '2' label twice, so that if a '2' is recorded at time t, the event is replaced by a '0' at time t, and a '1' at time t. Another method we tried was to split the event so that it occurs at time $t - .00001$ as a '0' and at time $t + .00001$ as a '1', where t is the time at which a '2' is recorded. However, the two methods yield identical results, so we use the former one.

We make the simplifying assumption that the distribution of the reaction times from white and black flashes are identical, so that they have the same parameters μ, σ. Recall that the location of the peaks for the nonparametric intensity estimates of $p_{S_W R_W}(u)$ and $p_{S_B R_B}(u)$ should be near the mean reaction time from white and

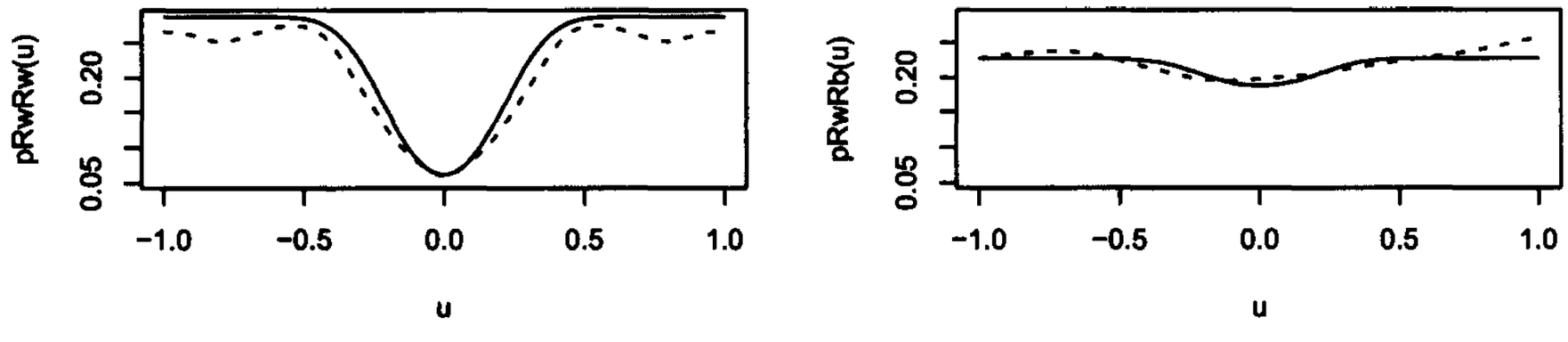


Figure 5.11: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for simulated data with $p_{S_W} = .7$, $p_{S_B} = .5$, and $d = .2$.

black flashes, respectively, because of the symmetric delay distribution. For data simulated from the model under the assumption of identical black and white reaction time distributions, the maximum difference observed between the peak locations of nonparametric $p_{S_W R_W}(u)$ and $p_{S_B R_B}(u)$ estimates was 0.05. Taking this observation into account, and based on plots of these nonparametric estimates for the nine pooled data sets, the assumption of equal means for the black and white reaction times appears to be reasonable for the majority of the data sets. Locations of the peaks for the nine pooled data sets are given in Table 5.5.

We also assume that the rates of the noise processes associated with the black and white flashes are both equal because of the equal stimulus rates, so that $p_{N_B} = p_{N_W} := p_N$. We use Algorithm 5.6.1 to obtain the estimates of μ, σ, and p_N. For the real data $p_{S_B} = p_{S_W} := p_S$, and we do not make the restricting assumption of $d = 0$. Thus, estimates of p, q, and d are found using the method of Section 5.6.2. In the parameter estimation $\hat{p}_{R_B R_B}(u)$ is employed since for each of the pooled data sets

Table 5.5: Peak locations for nonparametric estimates of $p_{S_W R_W}$ and $p_{S_B R_B}$ for pooled choice RT data, using a bandwidth of $h = .15$.

flash rate, p_S	$p_{S_W R_W}$ peak location	$p_{S_B R_B}$ peak location
0.2	.55	.54
0.3	.53	.54
0.4	.51	.56
0.5	.53	.53
0.6	.52	.55
0.7	.49	.56
1.0	.47	.49
2.0	.49	.52
4.0	.39	.48

the nonparametric estimates $\hat{p}_{R_B R_B,h}(u)$ exhibit the typical behavior for our model (a trough rather than a peak at 0), which is not always true for $\hat{p}_{R_W R_W,h}(u)$. The parameter estimates are given in Table 5.6.

Table 5.6: Parameter estimates for each of the nine pooled choice RT data sets

flash rate, p_S	$\hat{\mu}$	$\hat{\sigma}$	$\hat{d}$	$\hat{p}$	$\hat{q}$	$\hat{p_N}$
0.2	.551	.142	.025	.054	.033	.044
0.3	.522	.130	7×10^{-5}	.072	.054	.068
0.4	.511	.133	.079	.046	.046	.108
0.5	.504	.132	.007	.098	.028	.139
0.6	.504	.134	.084	0	.084	.188
0.7	.483	.121	.051	.110	.072	.227
1.0	.434	.109	.154	0	.128	.615
2.0	.465	.101	.099	0	.124	.761
4.0	.422	.065	.133	0	.261	.962

The mean reaction time, which is near 0.5, is higher than for the simple RT and go-no go RT experiments, where $\hat{\mu} \doteq 0.3$ and $\hat{\mu} \doteq 0.4$, respectively. This agrees with intuition since in a choice RT experiment more activities are required - a decision has to be made regarding the type of flash perceived, as well as which button to press. This result is consistent with previous observations (see Luce (1986)).

In general, the estimates of d are close to zero when both flash rates are .5 or less,

while $\hat{d} > 0$ for the faster stimulus rates. This result suggests that when flashes are presented at a high rate, temporal summation (two consecutive flashes viewed as one bright flash) occurs more frequently in the visual system. Recall that in our model, when two flashes are presented within d time units of each other they are perceived as one bright flash, and only one response occurs.

Completely random thinning does not appear to occur when the flash rate is high, as indicated by the $\hat{p} = 0$ for these cases. The error probabilities q give an indication of the proportion of wrong response types, and appear to increase with the flash rate. Such a result seems reasonable, since when the frequency of flashes is higher, mistakes are more likely to occur. As one may expect, the internal noise rate p_N appears to increase with the flash rate; when the stimuli are presented more frequently it may be more likely for a response to occur in the absence of a flash.

For simulations from our model $\hat{p}_{R_W R_W,h}(u)$ and/or $\hat{p}_{R_B R_B}(u)$ do not always have the expected trough at 0. Therefore, the absence of such a trough for our data is not necessarily an indication of our model's failure to capture the behavior for the choice RT data. Except for the poor fit for $p_{R_W R_W}(u)$, comparisons of the parametric and nonparametric estimates reveal that the correspondence between the two curves is closest for low to moderate flash rates (e.g. rates of .2 to .7 flashes/s for each flash type). Comparison plots of the intensity estimates for the nine choice RT data are given in Figures 5.12 to 5.28.

The estimate of d is negligibly different from zero for the data set with $p_S = .3$. Therefore, the response-response intensity estimates are constant and such plots are not considered.

In Figure 5.21 the difference in peak locations for $p_{S_W R_W}(u)$ and $p_{S_B R_B}(u)$ is clear, with the white response times having a slower rate. Despite the difference in peak locations, the similarities between the parametric and nonparametric estimates of $p_{S_B R_W}(u)$ and $p_{S_W R_B}(u)$ in Figure 5.21 suggest that our estimate of q is reasonable.

At faster stimulus rates ($p_S > 1.4$) our model does not fit well. A factor for this poor fit may be dependence between the reaction times. The plot of the $p_{S_B R_B}(u)$ and $p_{R_B R_B}(u)$ estimate in Figure 5.23 reveals a characteristic of a d value larger than

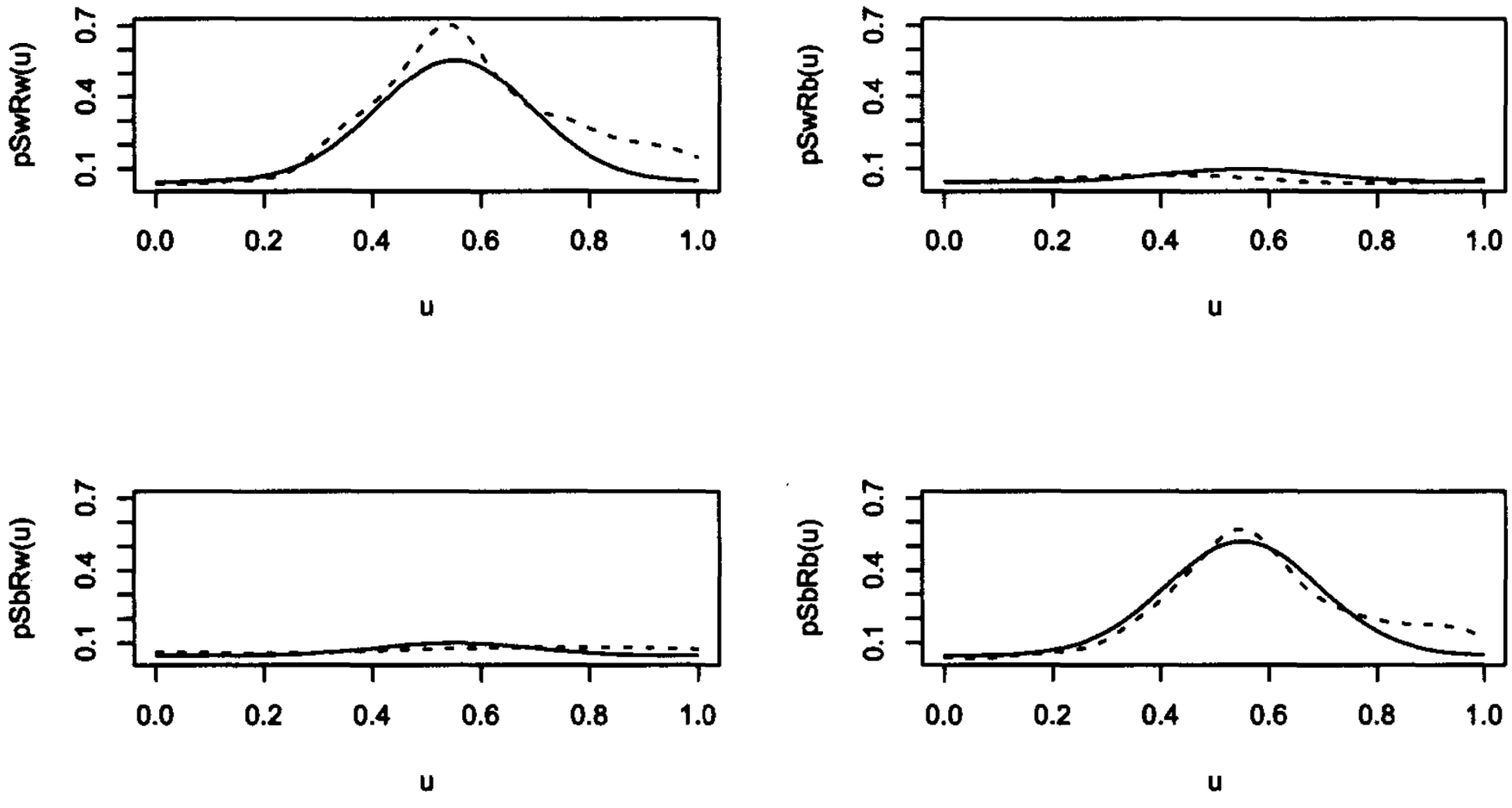


Figure 5.12: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .2$.

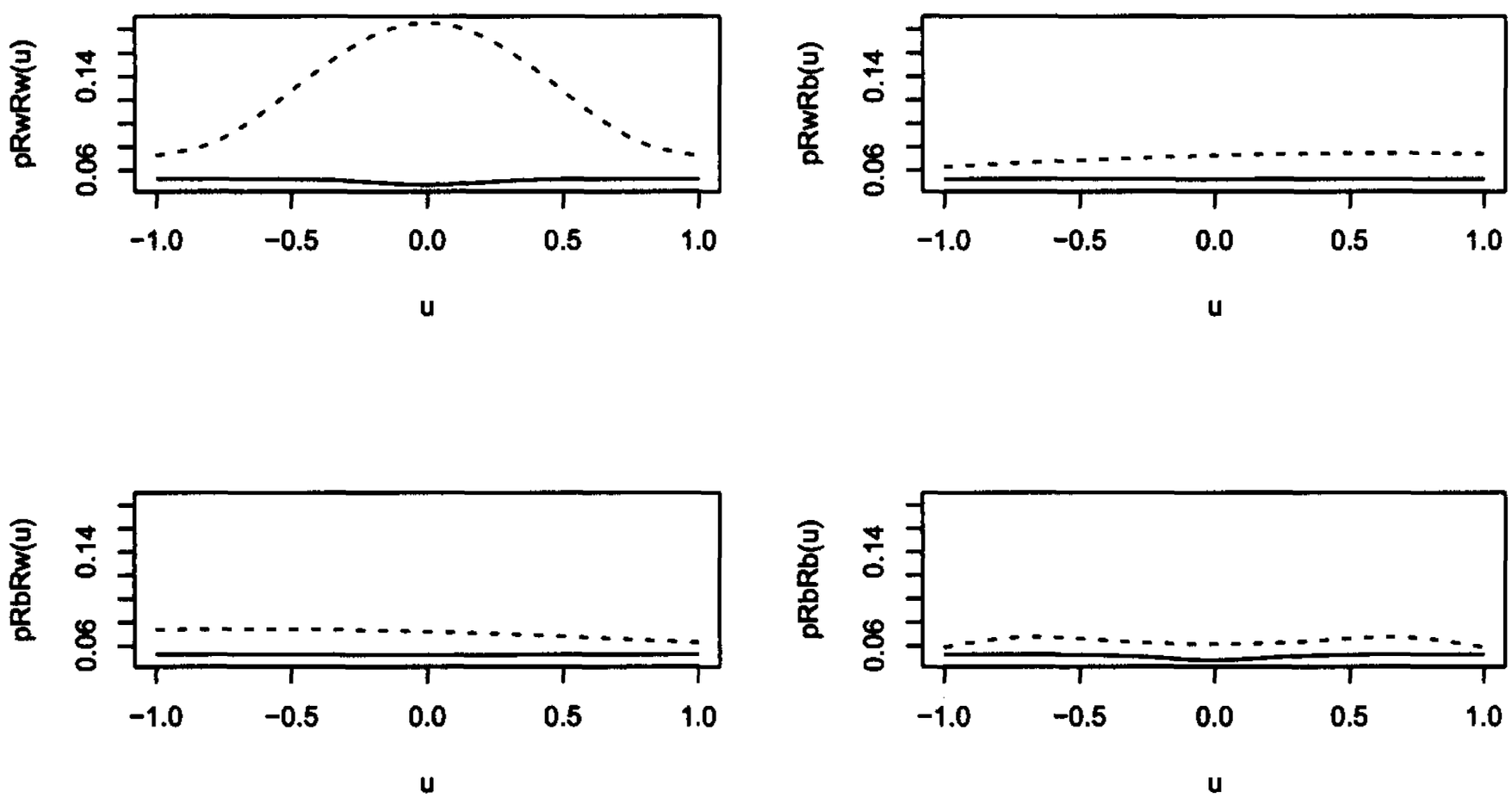


Figure 5.13: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .2$.

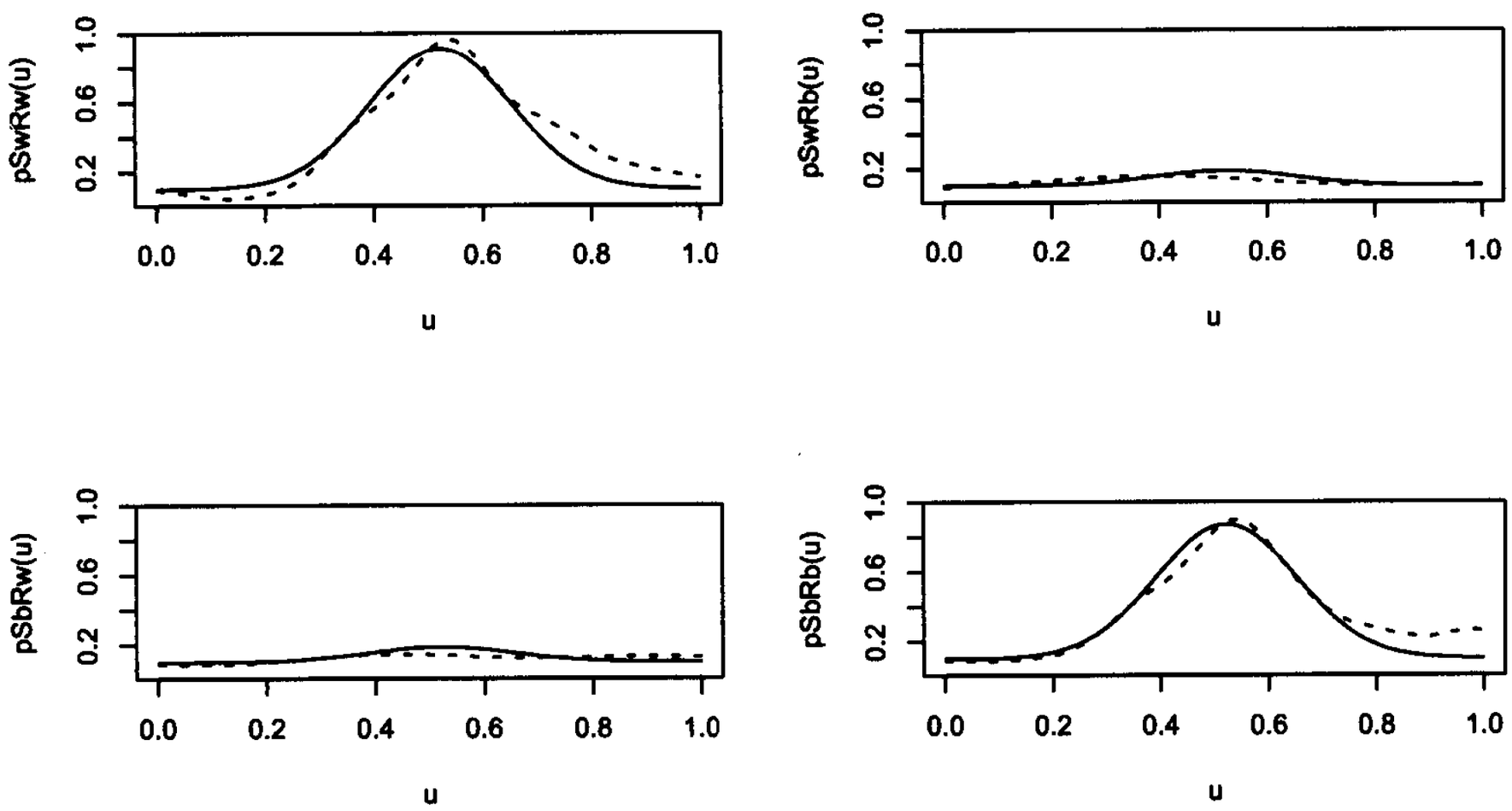


Figure 5.14: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .3$.

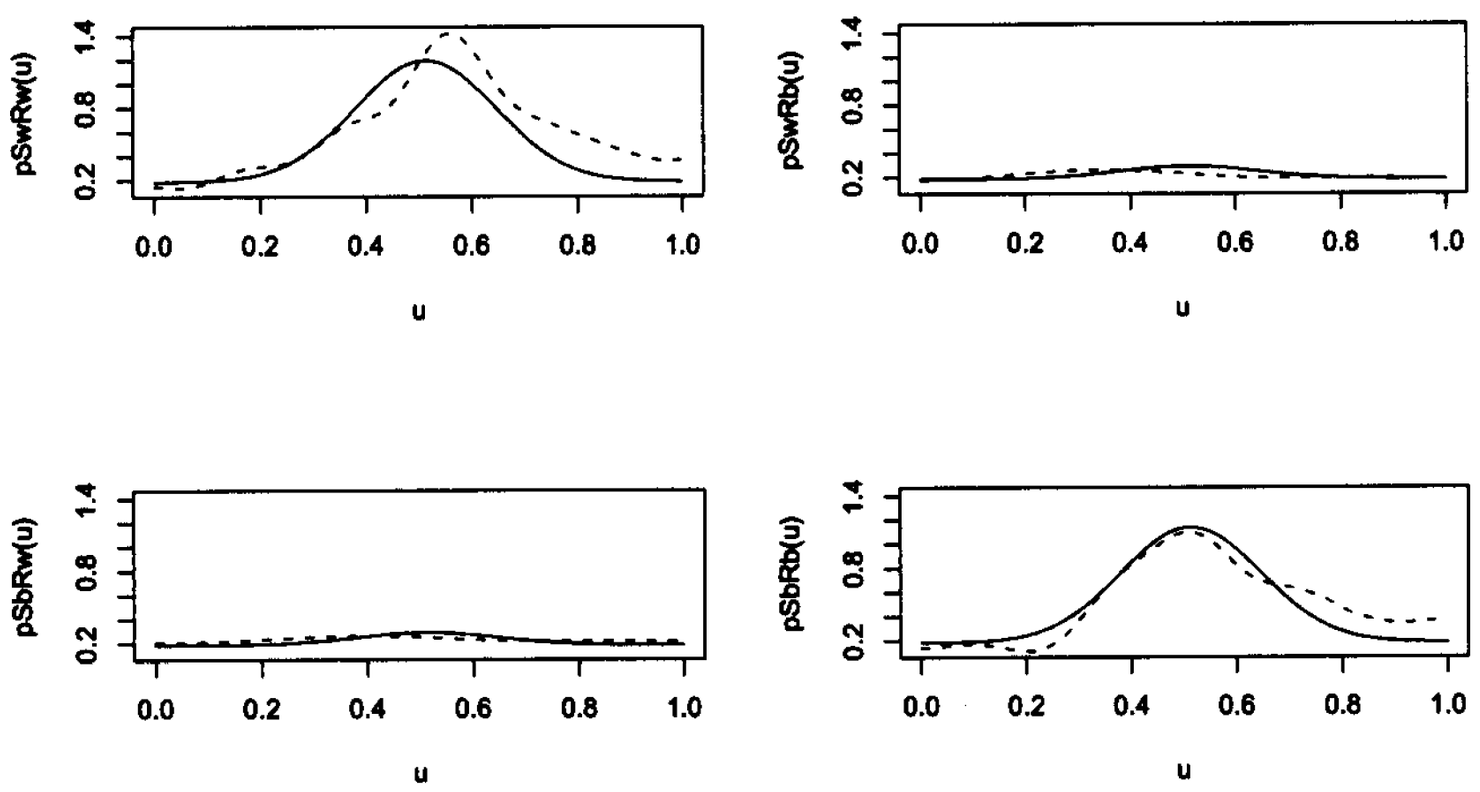


Figure 5.15: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .4$.

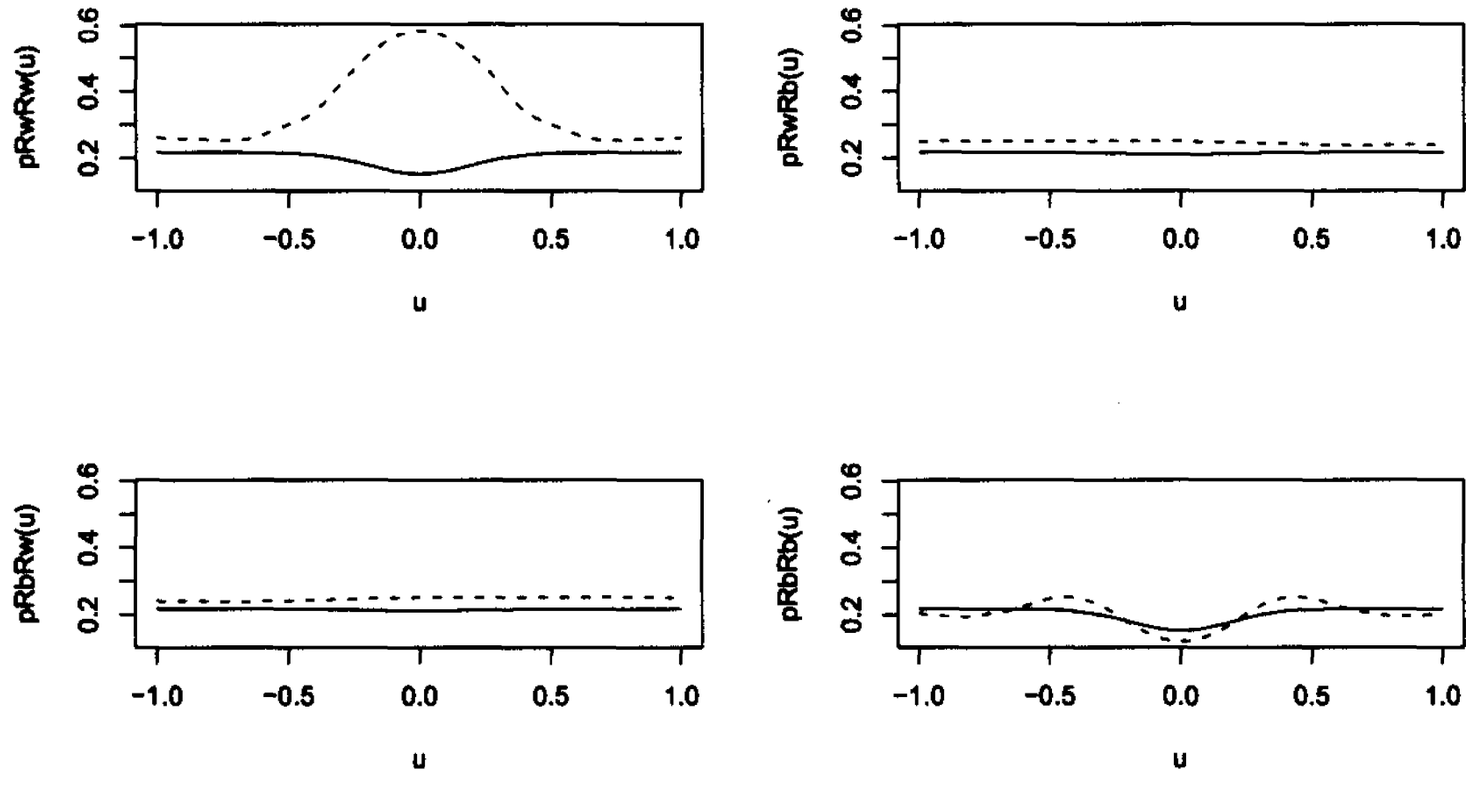


Figure 5.16: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .4$.

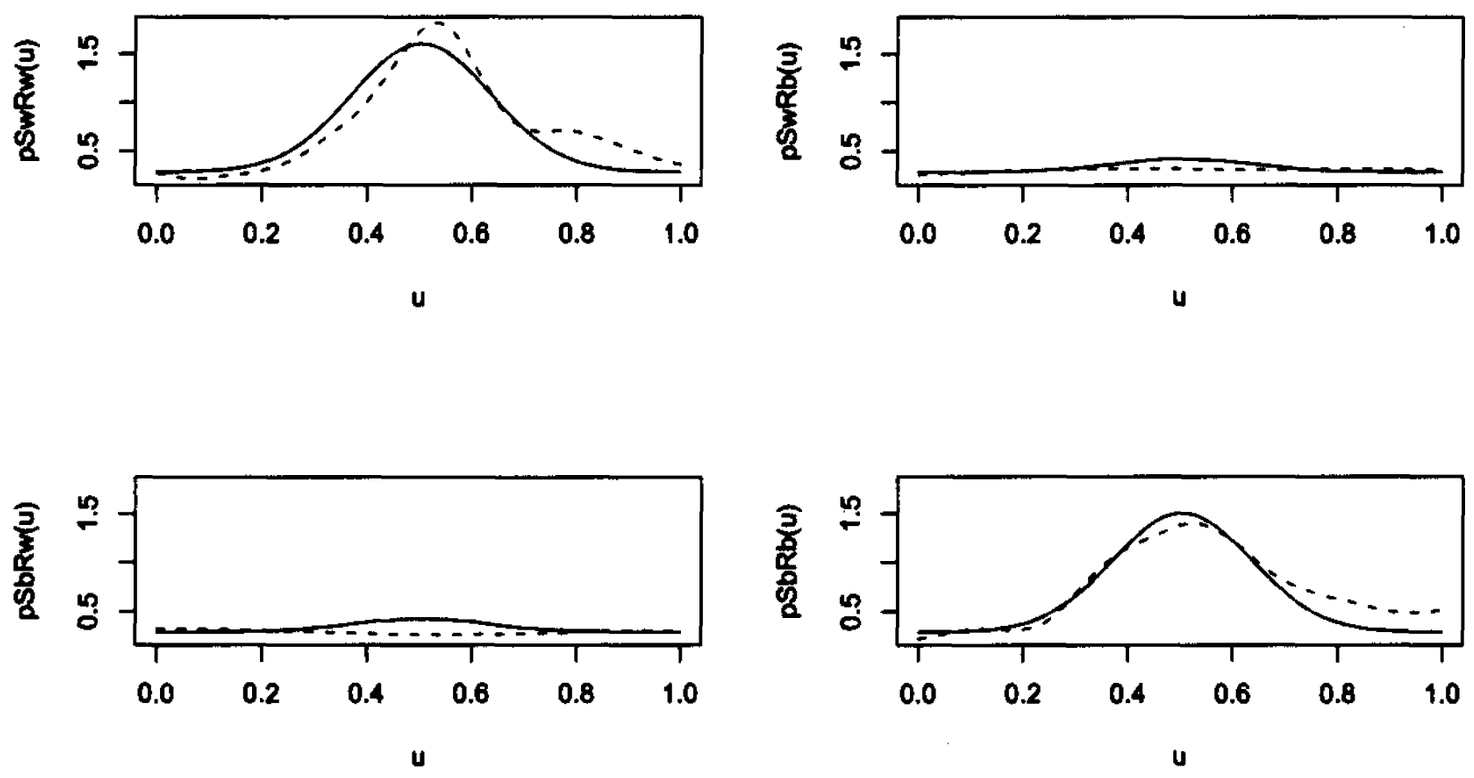


Figure 5.17: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .5$.

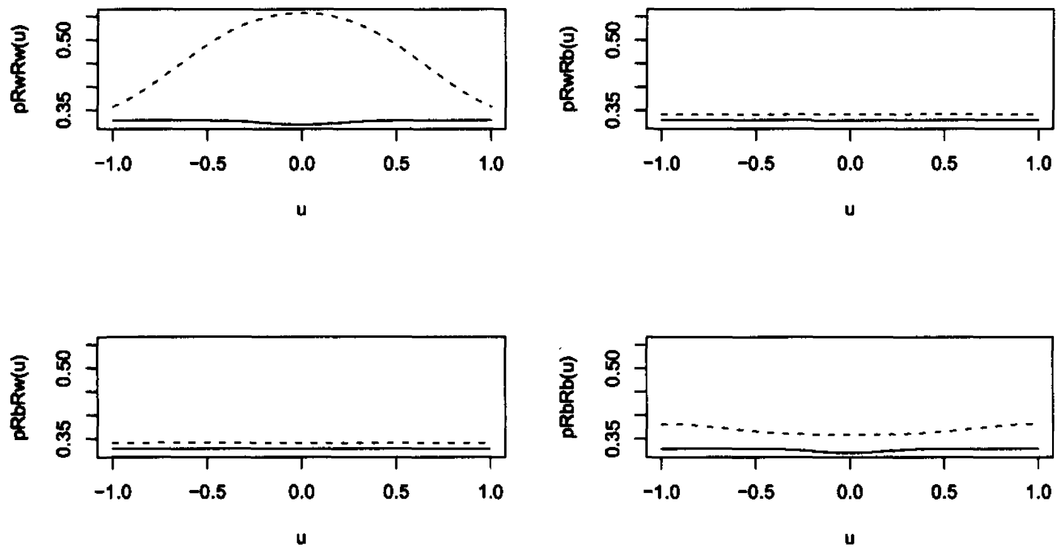


Figure 5.18: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .5$.

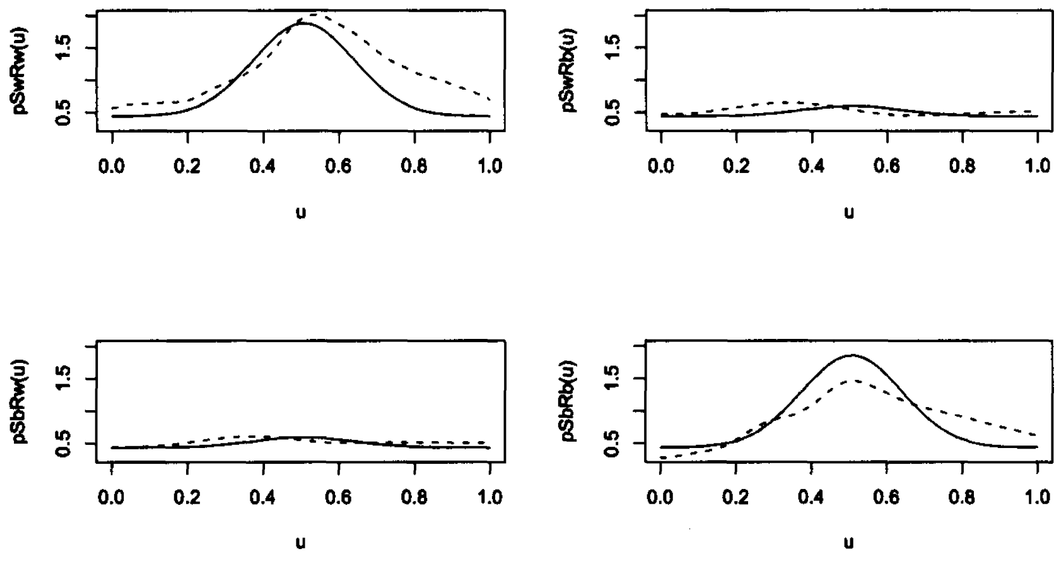


Figure 5.19: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .6$.

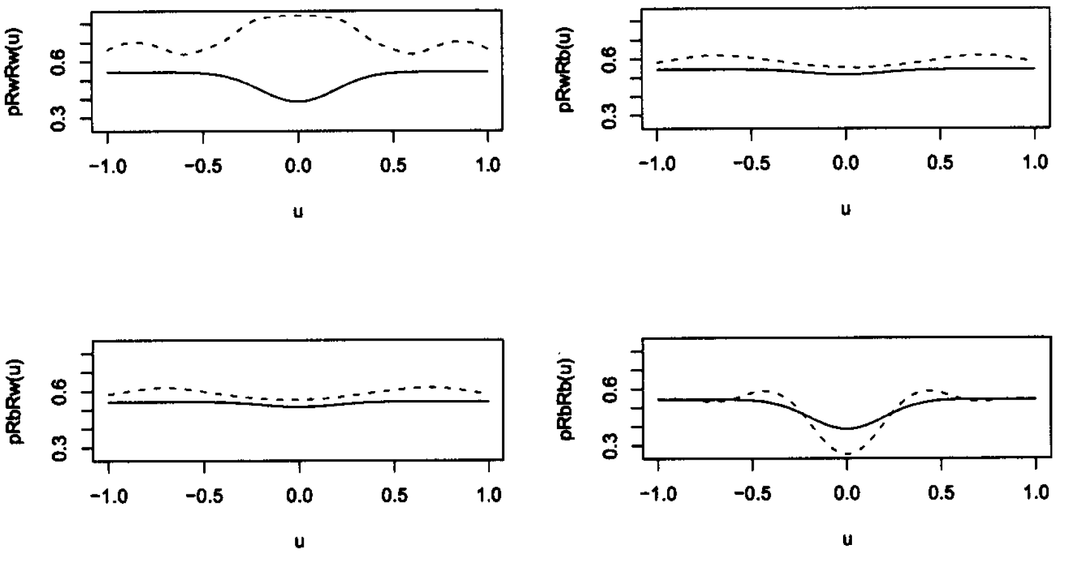


Figure 5.20: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .6$.

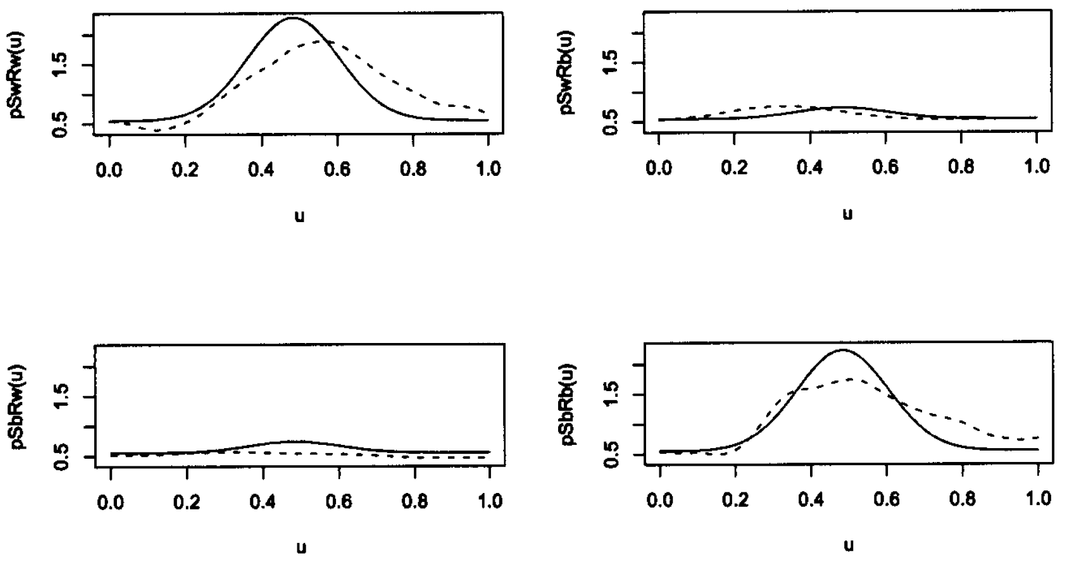


Figure 5.21: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .7$.

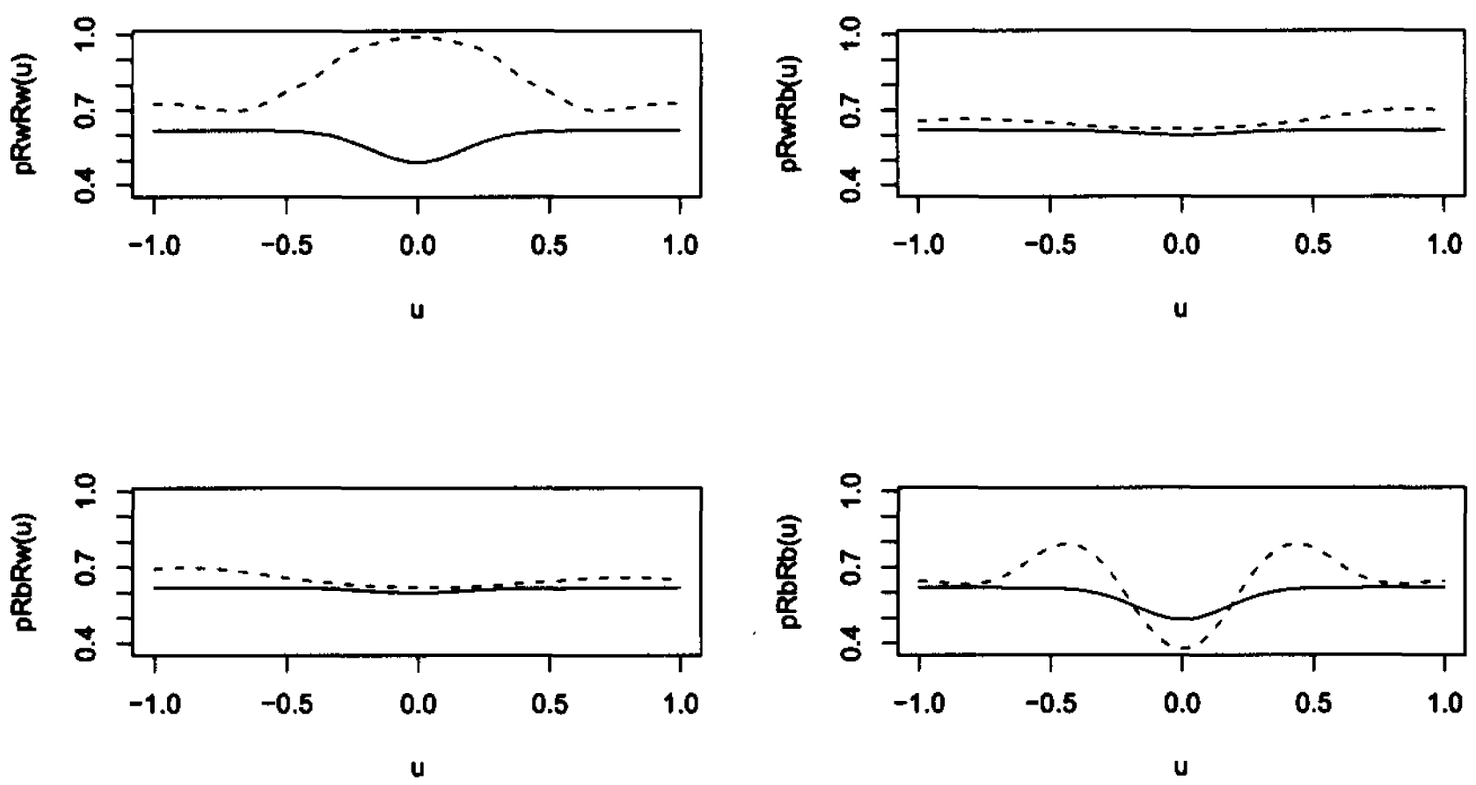


Figure 5.22: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = .7$.

what was estimated by the model; there is a depressed region to the left of the peak of $\hat{p}_{S_B R_B,h}(u)$. This suggests that there is more thinning in the data than predicted by our model. Possibly the extra thinning is because of dependencies between reaction times; at faster stimulus rates it is more likely that there is a time overlap between the time to respond to a flash, and the subsequent flash. Such an overlap is not likely as common in the slower flash rate data sets, which is probably why the model fits them better.

The two data sets with the fastest rates are the only ones in which all of the response-response intensities have a valley at lag zero. However, there is a poor correspondence between all of the nonparametric and parametric response-response intensities, as can be seen in Figures 5.26 and 5.28.

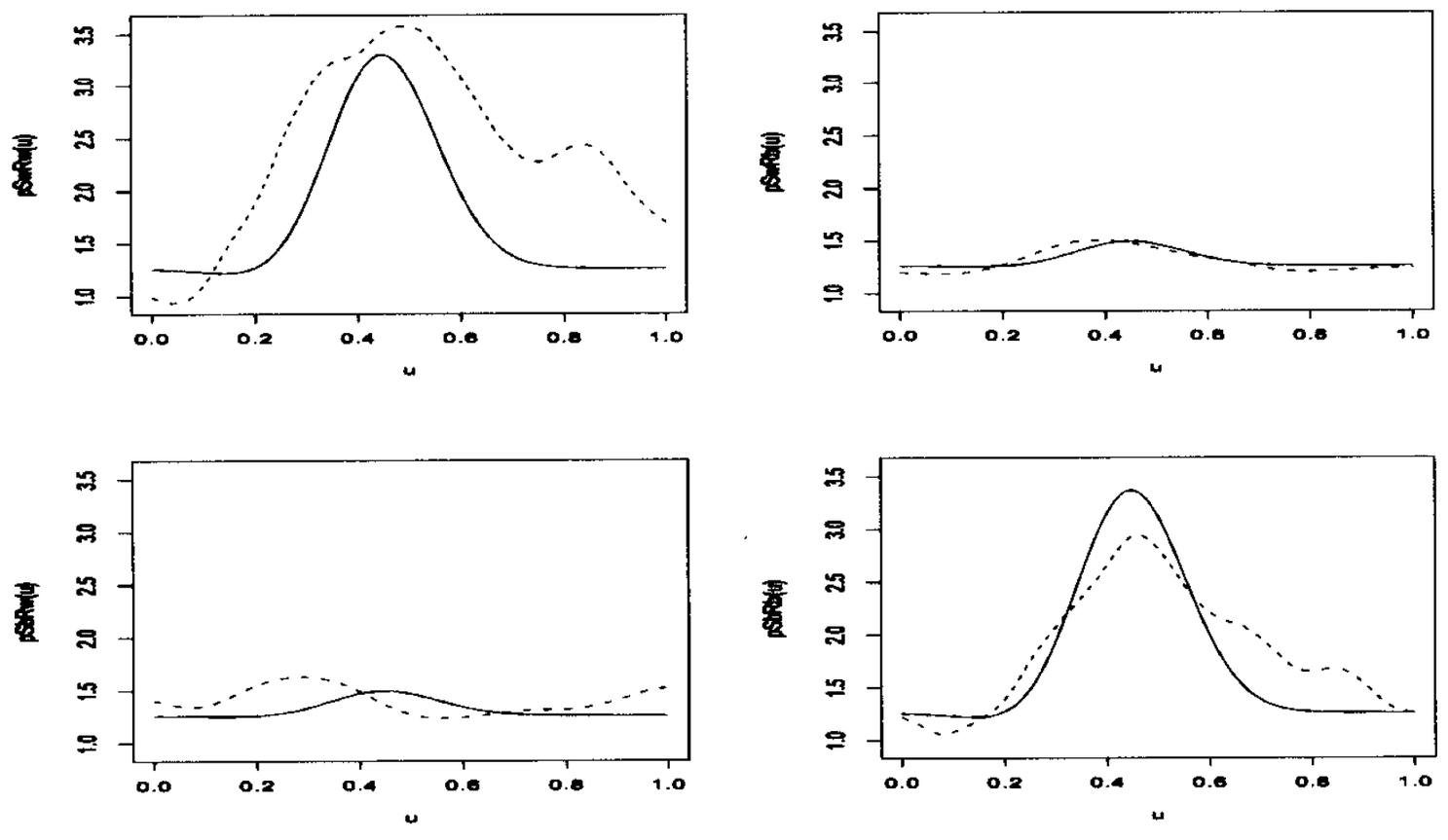


Figure 5.23: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = 1$.

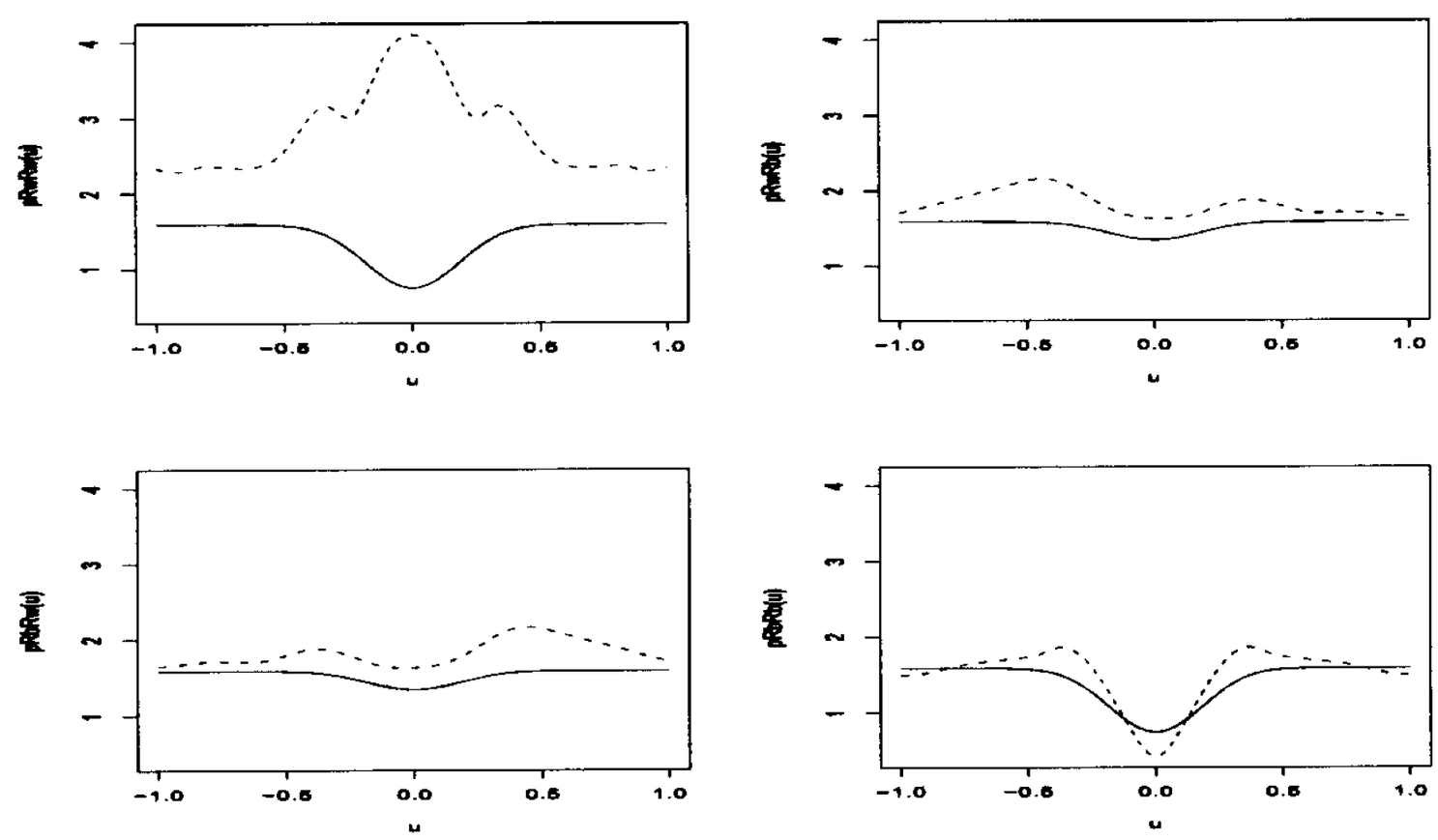


Figure 5.24: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = 1$.

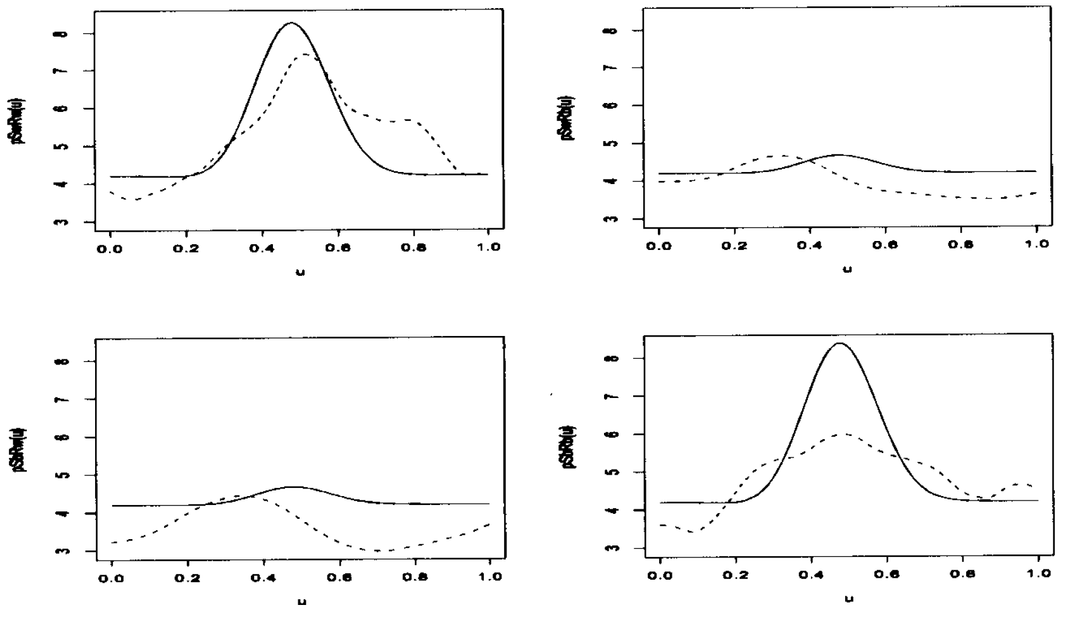


Figure 5.25: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = 2$.

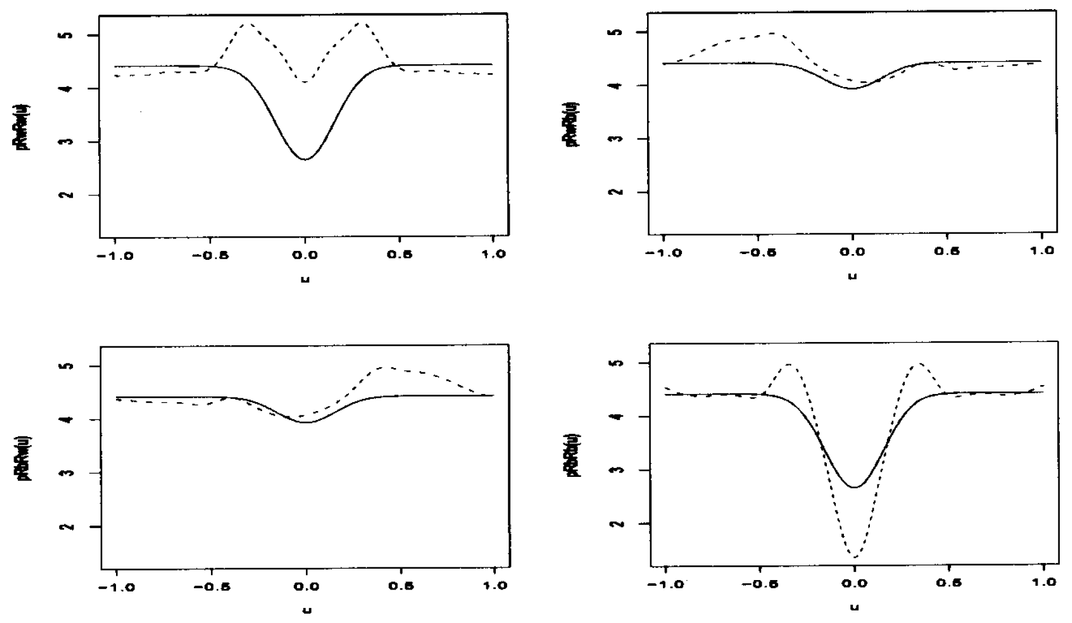


Figure 5.26: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = 2$.

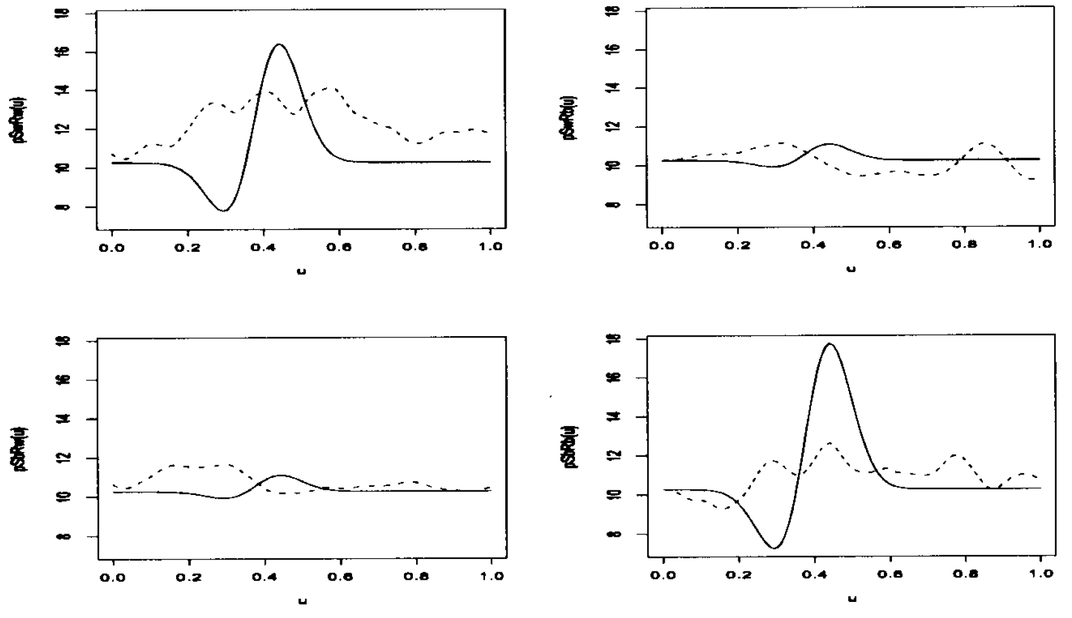


Figure 5.27: Second-order stimulus-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = 4$.

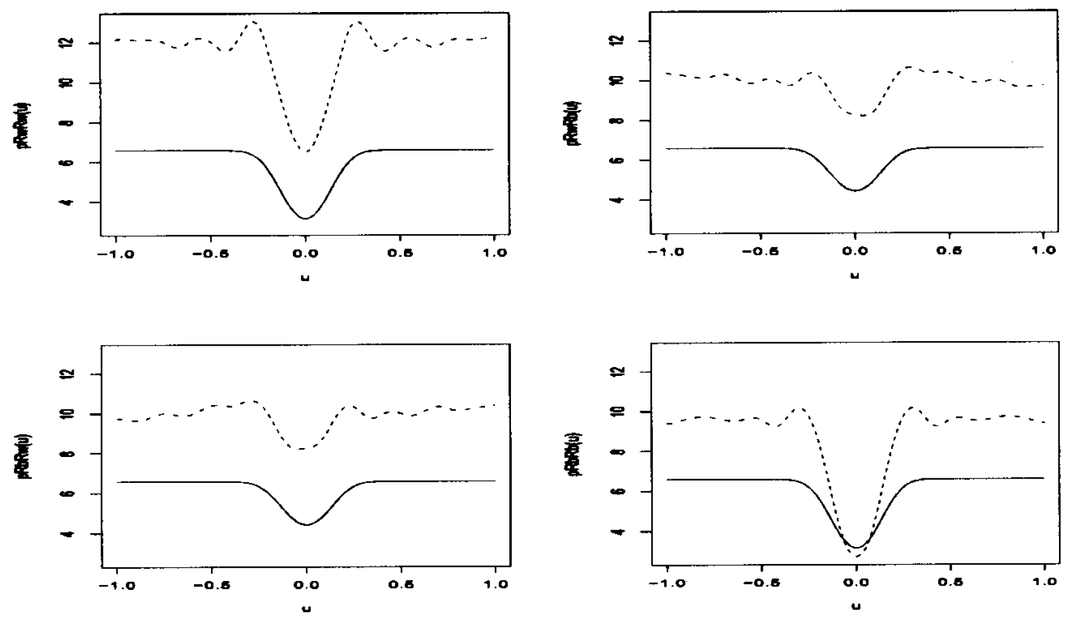


Figure 5.28: Second-order response-response parametric (solid lines) and nonparametric (dashed lines) intensity estimates for pooled choice RT data with $p_S = 4$.

As a comparison, we also studied parameter estimation for the unpooled data. The parameter estimates for the unpooled data tend to fluctuate around the estimates obtained for the pooled data. As an example, the parameter estimates for the 8 unpooled data sets with rate $p_S = .4$ have the means and variances as given in Table 5.7, where the pooled estimates are provided as well.

Table 5.7: Comparison of pooled and unpooled parameter estimates for choice RT data with flash rate $p_S = .4$

parameter	pooled estimate	mean of unpooled estimates	standard deviation of unpooled estimates
μ	.511	.505	.046
σ	.133	.126	.020
d	.079	.060	.064
q	.046	.059	.083
p	.046	.066	.062
p_N	.108	.109	.041

The standard deviations of the 8 unpooled estimates are quite small for each of the parameter estimates, and the means of the estimates are close to the estimates obtained for the pooled data. This gives support for studying the pooled data rather than the individual experimental runs. As a comparison, we estimate the means and standard deviations of the parameter estimates for 8 simulations with parameters equal to the pooled estimates for the data set with rate $p_S = .4$. Table 5.8 displays the results based on 500 samples of 8 simulations. There, it can be seen that the estimates have means that are quite close to the true values used in the simulations (except in the case of p), and small standard errors. Furthermore, the standard errors are quite similar to the empirically observed standard errors (given in Table 5.7), lending some further credibility to the parametric model.

5.9 Discussion

In this chapter we introduced a parametric model for choice RT data and developed methods for fitting such a model. Certain point process intensity functions were

Table 5.8: Standard errors of parameter estimates for 8 simulated data sets of 100 flashes with flash rate $p_S = .4$, based on 500 samples

parameter	true value	mean	standard error
μ	.511	.509	.022
σ	.133	.138	.047
d	.079	.059	.059
q	.046	.043	.049
p	.046	.132	.070
p_N	.108	.111	.020

studied in order to fit the model, and we derived intensity expressions under our model, as well as discussed nonparametric intensity estimation, including optimal bandwidth selection.

In the estimation of the parameters we made certain assumptions. We assumed that the reaction times from both black and white flashes are independent and identically distributed as $\mathrm{N}(\mu, \sigma)$, and that the noise processes associated with the black and white flashes have rates proportional to the respective flash process. Parameter estimation methods were developed under these assumptions, and illustrated on simulated data. Finally, we fit our model to nine choice RT data sets by applying our estimation methods. For the majority of the data sets the assumption of identical distributions for reaction times from black and white flashes appears to be reasonable.

Chapter 6

Threshold Models for a Choice RT Experiment

In a choice reaction time experiment two types of flashes, black and white, are presented against a grey background, to an observer. Upon perception of a flash the subject immediately presses one of two buttons, depending on the type of flash perceived, so that there are two response types.

In this chapter we extend the single response type threshold models studied in chapters 3 (simple RT) and 4 (go-no go RT) to two-response type threshold models. Fitting such models gives us a procedure for estimating both the black and white thresholds. Knowledge of the threshold values aids in making inferences regarding the minimum rod current level for each flash type to be detected.

In Section 6.1 we develop a threshold model with one threshold for each stimulus type, and a single internal potential for the two response types. Such a threshold model accounts for all of the information given in the data, but it is difficult to make inferences regarding the correct and error responses. This model is then extended further in Section 6.2, where we develop a marginal model for each response type, with a threshold for each marginal model. These marginal models are found to be more useful in making inferences on the intensities with which correct and error responses are made. Within this section we also extend our marginal models to incorporate bivariate interaction terms, which allow us to study interactions between two flashes

at different distances from a single response. For each of the models considered we discuss the statistical methods involved and diagnostics, as well as fit the models to both simulated data based on the parametric model of Chapter 5, and real data. We conclude this chapter with a discussion in Section 6.3.

6.1 A Threshold Model for Choice RT Data

In this section, we extend the model used for a single response type to data with two response types, and two thresholds - one threshold for each response type. Prior to describing this threshold model, we describe our conceptual model for a choice RT experiment.

In our conceptual model, the stimuli, which consist of both black and white flashes, pass through an internal filter. We set up the model such that black flashes give negative contributions, while white flash contributions to the filter are positive valued. We model the internal potential as the filter output with internal noise added to it. When the internal potential drops below the 'black' threshold θ_B the subject decides that there is a black flash and has a 'black' response. A 'white' response occurs when the internal potential exceeds the white threshold θ_W.

If the subject tends to make each response type with the same frequency when both flash types are equiprobable, then the subject has no bias in responding to the two flash types, and $\theta_B = -\theta_W$.

As in previous chapters, we discretize a point process S into the 0-1 time series X as follows, where we choose some lag value m to create the time intervals $\{(0, m], (m, 2m], \ldots, (T, T+m]\} = \{I_1, I_2, \ldots, I_T\}$, where $T = m\lceil \max(R^B, R^W)/m \rceil$. The lag m may be chosen such that there is at most one flash (response) in each discretization interval, but to decrease computation time we choose m slightly larger, allowing more than one point in a given interval.

We use the following notation:

- S^B and S^W denote the black and white stimulus point processes, respectively;

- R^B and R^W denote the black and white response point processes, respectively;
- X^B and X^W denote the discretized black and white stimulus point processes, respectively;

$$X_j = \begin{cases} 1, & \text{if there exists } S_k \text{ such that } S_k \in I_j \\ 0, & \text{otherwise} \end{cases}$$

- Y denotes the discretized response process, and takes on the values $\{-1, 0, 1\}$ as follows:

$$Y_j = \begin{cases} -1, & \text{if there exists } R_j^B \text{ such that } R_j^B \in I_j \\ 1, & \text{if there exists } R_j^W \text{ such that } R_j^W \in I_j \\ 0, & \text{otherwise} \end{cases}$$

Occasionally $Y_t = 2$ due to a black and white response occurring almost simultaneously. For each such case, with equal probabilities we set $Y_t = 1, Y_{t+1} = -1$, or $Y_t = -1, Y_{t+1} = 1$. The same results are obtained when the simpler assignment of setting $Y_t = 1, Y_{t+1} = -1$ for all case of $Y_t = 2$.

We denote the internal potential at time t by U_t. Possible models for U_t are considered later in this section. In our model, a black response occurs when

$$U_t + \varepsilon_t < \theta_B,$$

and a white response occurs when

$$U_t + \varepsilon_t > \theta_W,$$

where ε_t is internal noise, which we assume to have a symmetric distribution about 0. We assume a logistic error distribution.

This set-up corresponds to an ordinal logit regression model (Skrondal and Rabe-Hesketh (2004)). The observed response y_t takes on one of three response values $\{-1, 0, 1\}$, while the latent response has the form $U_t + \varepsilon_t$, where ε_t is an error term, and the relationship between the observed and latent responses is given by

$$y_t = \begin{cases} -1, & \text{if } U_t + \varepsilon_t < \theta_B \\ 0, & \text{if } \theta_B < U_t + \varepsilon_t < \theta_W \\ 1, & \text{if } U_t + \varepsilon_t > \theta_W \end{cases} .$$

We will use the same notation as in the single response cases of simple and go-no go RT, but we describe it here for convenience. Assuming that a flash occurs at time τ, $a_{t-\tau}$ represents the effect of that flash on the internal potential at time t. The summation function a_u acts as a filter on the flash process. Note that time is in terms of an index corresponding to the interval containing the time of interest.

As of time t, the number of time intervals elapsed since the last time interval containing a response is denoted by γ_t. We assume that a response occurs at time 0. Thus, $\gamma_1 = 1$, and letting k be the interval in which the first response occurs, we have $\gamma_j = j$ for $j = 1, \ldots, k$. In general, $\gamma_j = j - i$ for $j = 1, \ldots, T$, where the i th interval I_i, was the last interval before I_j to contain a response.

We start with a linear model as an approximation to the internal potential. Let a_u and b_u be the summation functions for white flashes and black flashes, respectively. Under this assumption, the internal potential can be expressed as follows (Brillinger (1988b))

$$U_t = \sum_{u=0}^{g_t-1} (a_u X_{t-u}^W + b_u X_{t-u}^B), \tag{6.1.1}$$

where $g_t = \min\{\gamma_t, G\}$ and G is the index of the corresponding interval that contains the maximum time lag of interest. We truncate the sum at G if $\gamma_t > G$ since based on physical grounds, there is no reason to believe that effects continue past interval G.

In the simplest case $b_u = -a_u$, and the model is

$$U_t = \sum_{u=0}^{g_t-1} a_u (X_{t-u}^W - X_{t-u}^B).$$

This corresponds to the case in which black and white flashes lead to effects of the same magnitude on the internal potential in a tendency towards a response of the same type as the flash; the effect of a black flash in decreasing U_t is the same as the effect of a white flash in increasing U_t. Intuitively, this appears to be a reasonable assumption. Note that the assumption of effects of equal magnitude does not imply no response bias. That is, there is no restriction that $\theta_B = -\theta_W$; it is possible that when $U_t = C > 0$ we have $U_t > \theta_W$, but when $U_t = -C < 0$ we do not have $U_t < \theta_B$.

Thus, we make the assumption of equal effects $b_u = -a_u$, and focus on this simplest model, including the term $X_t^D = X_t^W - X_t^B$ rather than the individual X_t^W and X_t^B, so that our model becomes

$$U_t = \sum_{u=0}^{g_t-1} a_u X_{t-u}^D.$$

We next look at an extension of the linear model that includes interaction terms. Inclusion of such terms takes into account the fact that a response to a flash may be dependent on the occurrence of previous flashes that are close together in time with the present one. That is, we consider interactions of a flash at time t with the occurrence of a flash of a certain type that is up to d time units earlier, for some d. For simplicity we set $d = rm$, for some integer r, since we are working with data that is discretized using lag m. Letting $Z_j^B = I\{\sum_{i=1}^r X_{j-i}^B > 0\}$, and $Z_j^W = I\{\sum_{i=1}^r X_{j-i}^W > 0\}$ the internal potential can be expressed as

$$U_t = \sum_{u=0}^{g_t-1} a_u X_{t-u}^D + \sum_{u=0}^{z_t^B-1} b_u X_{t-u}^D Z_{t-u}^B + \sum_{u=0}^{z_t^W-1} c_u X_{t-u}^D Z_{t-u}^W, \tag{6.1.2}$$

where $z_t^B = \min\{\gamma_t, z^B\}$, $z_t^W = \min\{\gamma_t, z^W\}$, and $z^B, z^W \leq G$ are the indices of the interval containing the maximum interaction time lag of interest, involving black and white flashes, respectively. In some cases, there may be two different d parameters, d^B for black stimuli and d^W for white stimuli.

Interaction effects give an indication of the presence of inhibition or facilitation among the flashes. For example, negative (positive) b_u points toward inhibition (facilitation) between black flashes, since $X_{t-u}^D Z_{t-u}^B$ is negative when there are two black flashes within d^B time units. That is, two black flashes within d^B time units will give a positive contribution to U_t when $b_u < 0$ so that a response by crossing the lower threshold is less likely to occur, indicative of inhibition among the black flashes. Likewise, facilitation among the black flashes is present if $b_u > 0$, since the contribution of two black flashes within d^B time units will be negative, giving a negative contribution to U_t and crossing the lower threshold is more probable. It should be noted that negative (positive) b_u also signifies an inhibition (facilitation) effect on a white flash that follows within d^B time units of a black flash. Similar reasoning can be used to

show that negative (positive) c_u points toward inhibition (facilitation) between white flashes, and to a black flash following a white flash.

We fit ordinal logit regression models to the data in R by using the `polr` function in the **MASS** library (Ripley (2006)). Our fitted model with interactions has the form:

$$\hat{Y}_t = \begin{cases} -1, & \text{if } \hat{U}_t < \hat{\theta}_B \\ 0, & \text{if } \hat{\theta_B} < \hat{U}_t < \hat{\theta}_W \\ 1, & \text{if } \hat{U}_t > \hat{\theta}_W \end{cases},$$

where

$$\hat{U}_t = \sum_{u=0}^{g_t-1} \hat{a}_u X^D_{t-u} + \sum_{u=0}^{z^B_t-1} \hat{b}_u X^D_{t-u} Z^B_{t-u} + \sum_{u=0}^{z^W_t-1} \hat{c}_u X^D_{t-u} Z^W_{t-u}.$$

Models with interactions are fit to nine pooled data sets (pooling based on 8 runs of 100 flashes), each having a different stimulus rate. We pool the data as described in Section 3.3.6. The rates of flashes per second are 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 2.0, 4.0, and 8.0. Black and white flashes are presented with equal rates, so that if the stimulus rate is 1.0, each flash type is presented with rate $p_S = .5$.

We can assess goodness of fit of the model by examining plots of the empirical probabilities of a black response ($Y = -1$) and a white response ($Y = 1$) against the predictor $\hat{U}_t$. The probability of a white (black) response will increase (decrease) with U_t since a response occurs when the internal potential, with some random noise added to it, crosses the white threshold θ_W (black threshold θ_B); the larger the absolute value of U_t, the higher the probability of some type of response.

Using a grid of u values that span the range of the $\hat{U}_t$ the empirical probability of a white (black) response is the proportion of white (black) responses that occur when $\hat{U}_t$ is in a small interval near each u. An estimate of the probability of a white response is given by

$$\frac{\sharp\{Y_t = 1 \text{ with } u - h < \hat{U}_t < u + h\}}{\sharp\{t \text{ with } u - h < \hat{U}_t < u + h\}},$$

where h is taken to be a small positive value, and the empirical probability of a black response is found by replacing $Y_t = 1$ by $Y_t = -1$ in the above ratio. This estimate of the probability is not very sensitive to h; the general pattern of the probability

estimates is similar for different h. However, for $h < .15$ there tends to be more estimates equal to one, due to the small interval width $2h$. We use $h = .2$ in the probability estimation.

Under the assumption that the model is correct, the probability of a white response is

$$P(U_t + \varepsilon_t > \theta_W) \doteq P(\varepsilon_t > \hat{\theta}_W - \hat{U}_t),$$

and the probability of a black response is

$$P(U_t + \varepsilon_t < \theta_B) \doteq P(\varepsilon_t < \hat{\theta}_B - \hat{U}_t).$$

The plots of the two empirical probabilities should follow the corresponding theoretical curve if the model fits the data well.

Before fitting our model to real data, we first fit our model to simulated data from our parametric model in Chapter 5, in order to study the behaviour of the fitted model and diagnostic for data in which we know the relationship between the flashes and responses. The data is simulated using Algorithm 5.3.1.

Plots of the linear filter estimates exhibit similar behavior regardless of the stimulus rates, but the peak height tends to increase with flash rate. As q increases to 0.5 the linear filter tends to fluctuate around zero, and most coefficients are near zero. For some of the simulated data sets, as d increases a region to the left of the peak is depressed. This behaviour can be seen in Figure 6.1 for the two curves corresponding to simulated data with $p_S = .5$ and $d \in \{0, .1\}$.

Interaction effects that improve the model tend to be negative, suggesting that there are inhibition effects among flashes of the same type. This is expected since the model from which our simulations are generated incorporates deletions of flashes of the same type that are "close" in time, when $d > 0$.

Examples of threshold value estimates for various simulated data sets are provided in Table 6.1. For the simulated data we know that there is no response bias, so $\theta_B = \theta_W$, and we expect the two threshold estimates to be approximately equal in magnitude. At faster flash rates, regardless of the parameter values, there is little to no difference in magnitude between $\hat{\theta}_B$ and $\hat{\theta}_W$. However, at slower flash rates

for realistic (small) values of q, the differences in magnitude between $\hat{\theta}_B$ and $\hat{\theta}_W$ are slightly larger; as q approaches 0.5, the two estimates are similar in absolute value.

For a given flash rate, the threshold estimates tend to be very similar across various parameter settings. Furthermore, the threshold estimates tend to decrease in magnitude as the flash rate increases. The behavior of decreasing threshold magnitudes with increasing flash rates is expected since the response frequency should increase as the flash rate increases. The internal potential U_t will cross the thresholds more frequently when there is a fast stimulus rate; in a short time interval more flashes occur so that after a small amount of time a response occurs. In order for the internal potential to cross the thresholds within a short period of time it is anticipated that the magnitude of the thresholds must not be large when the flash rate is high. Furthermore, such results are consistent with the observation that the peak heights of the linear filter decrease as the stimulus rate increases; if the threshold is low then the coefficients in the linear filter do not need to be large for a response to occur.

Table 6.1: Examples of threshold estimates for simulated data. The upper value in each table element is $\hat{\theta}_B$, while the lower term is $\hat{\theta}_W$.

	(a) $p_S = .5$		(b) $p_S = 2$	
$d \backslash q$	.05	.5	.05	.5
0	-4.58	-3.87	-2.54	-2.46
	4.17	3.81	2.54	2.55
.1	-4.59	-3.81	-2.83	-2.66
	4.28	3.94	2.70	2.67
.2	-4.57	-3.98	-2.90	-2.74
	4.27	3.98	2.87	2.83

The diagnostics of the fitted models for the simulated data tend to improve as the stimulus rates increase and as the thinning parameter d of the parametric model increases. When nonlinear inhibition is absent in the parametric model ($d = 0$), the diagnostics indicate that the fit is relatively good, but there is a tendency for the threshold model to overestimate the probability of a response at large magnitude U_t. As the thinning parameter increases to 0.1 and to 0.2, the fits improve, and for faster rates such as $p_S = 2.0$ the fit becomes very good.

The parametric model can assist us in interpreting the behaviour in our diagnostic plots. In the parametric RT model many consecutive non-responses to flashes do not increase the probability of a response, but in the threshold model the probability increases with the number of consecutive non-responses. However, in our parametric model, when $d = 0$ there are fewer instances in which a flash is not responded to, and as d increases the number of non-responses increases. Thus, the improvement in fit of the threshold model to simulated data with larger d settings may be because with larger d values, non-responses occur more frequently in the simulated data. That is, in our parametric model more non-responses may give the appearance that the probability of a response increases with the number of non-responses, which is true for the threshold model, thus providing a good fit.

An example of the diagnostic for a simulation with parameters $\mu = .5$, $\sigma = .12$, $d = 0$, $p = .1$, $p_{S_B} = p_{S_W} = .5$,and $q = .05$ is given in Figure 6.2, while an example of the diagnostic for a simulation with parameters $\mu = .5$, $\sigma = .12$, $d = .2$, $p = .1$, $p_{S_B} = p_{S_W} = 2$,and $q = .05$ is given in Figure 6.3.

We now discuss the fitted threshold models for the experimental data. The "best-fitting" model, based on the diagnostic plots, for each of the pooled data sets has $d^B = d^W = d = .2$, but varies in the number of interaction terms.

A plot of the coefficients of lags of the flash process for each of the nine data sets is provided in Figure 6.4. It can be seen that the peak locations for each curve are all near a lag of 0.5 s, which indicates that most responses tend to occur 0.5 s after a flash, regardless of the flash type. This agrees with our estimates of the mean reaction time based on parametric models, in Chapter 5. The curves also appear to flatten out and have lower peak heights as the flash rate increases. The pattern of decreasing peak heights as the stimulus rate increases is evident in Figure 6.5. Such behaviour was also observed in the fits for simulated data.

The coefficients of the main effects included in the fitted models are given in Table 6.2, where a lag of $m = .05$ was used. For illustration purposes, in Figure 6.4 we included terms that are negligibly different from zero, but in Table 6.2 we only include those that are significantly different from zero. In the output from R a t-value

is provided for each coefficient. We set level of significance as $\alpha = .05$, so that those with a t-value larger in absolute value than 1.96 are included in the model.

Table 6.2: Main effect coefficients included in ordinal logistic threshold models fit to choice RT data.

Lag \ rate	0.4	0.6	0.8	1.0	1.2	1.4	2.0	4.0	8.0
0							-.26		
.05	-.48			-.34					
.1									.43
.15						-.28			-.34
.25								.41	
.3	.60	.84			.77	.56	.53		
.35	1.89	1.59	1.62	1.71	.22	.85	.40		
.4	2.69	2.16	1.71	1.67	.56	.83	.69		
.45	2.44	2.26	1.98	1.91	1.56	1.58	1.19	.45	
.5	3.69	3.21	2.66	2.73	1.48	1.64	.86	.79	
.55	3.93	3.23	2.54	2.48	1.79	1.37	.91	.78	.80
.6	3.84	3.01	2.51	2.49	1.72	1.64	1.09	.70	
.65	3.46	2.10	1.70	1.79	1.41	1.23	.81		
.7	2.70	2.54	2.00	1.39	1.49	.97	.80	.69	
.75	2.34	1.66	1.54	.91	.60	.84	.45		
.8	2.57	1.35	.99	.61	.54	.64	.40	.68	1.18
.85	1.70	1.14	.92	.53	.79	.56	.49		
.9	1.39	.82	.65	.40	.54				
.95	1.50	1.07		.39					
1.00		.92	.47						-1.07

Interaction effects that are included in the model tend to have negative coefficients, suggesting that there are inhibition effects among flashes of the same type. Both the black and white flash interaction coefficients included tend to be less than 0.2 lags above or below the filter peak height of 0.5. Recall that in our threshold model we used $d_B = d_W = .2$. This indicates that at lags near the mean reaction time inhibition effects occur frequently. That is, at lags near the mean reaction time, when two flashes of the same type are presented within 0.2 s apart the two flashes are perceived as one bright flash - temporal summation of the flashes occurs.

Tables 6.3 and 6.4 provide the interaction coefficient estimates at lags included in the fitted models of each data set.

Table 6.3: Black flash interactions included in ordinal logistic threshold models fit to choice RT data.

Lag \ rate	0.4	0.6	0.8	1.0	1.2	1.4	2.0	4.0	8.0
0		-1.47							
.15					-1.21	1.15			
.25	-1.35		-.92	-.87		-1.00	.94		
.3		-1.26			-.78				.60
.35	-2.93	-1.75	-1.92	-1.84	-1.07		-1.05		
.4		-2.05	-1.91	-1.34		-1.04	-.60		
.45	-2.11	-2.82	-2.04	-1.27		-1.34			
.5	-2.38	-1.61		-1.83	-1.08			-.94	
.55	-3.77	-2.39		-.86		-.96			
.6	-1.98								
.65				-1.15	.85				
.7		-2.49		-1.62	-2.11	-1.26		-.76	-1.49
.75				.9					
.8				-1.54				-.98	

The two threshold estimates, $\hat{\theta}_B, \hat{\theta}_W$, for our fitted models are provided in Table 6.5. The absolute values of the threshold estimates tend to decrease with the stimulus rate, which is the same behaviour observed for simulated data.

For a given total flash rate of at least 1.4 s, there are small differences in the black and white threshold magnitudes, so that the magnitude differences between $\hat{\theta}_B$ and $\hat{\theta}_W$ are negligible. This relation suggests that there is negligible bias in response choice when flashes are presented at moderately high speeds. The difference in magnitude is slightly larger for slower flash rates. For the simulated data we observed that for realistic values of q the differences in magnitude between $\hat{\theta}_B$ and $\hat{\theta}_W$ are larger when flashes are presented at a slower rate. Therefore, these results suggest that at all flash rates, there is negligible bias in response choice.

As an example of the diagnostic results, diagnostic plots are provided for four of the nine data sets in Figure 6.6. As observed for the simulated data with slow stimulus rates, the diagnostic plot for the data set with the slowest rate, 0.4, reveals a poor fit for predicting the probability of a black response when $\hat{U}_t < \hat{\theta}_B$; the probability of a black response is overestimated by the threshold model. The fitted and theoretical

Table 6.4: White flash interactions included in ordinal logistic threshold models fit to choice RT data.

Lag \ rate	0.4	0.6	0.8	1.0	1.2	1.4	2.0	4.0	8.0
.05								-.34	
.1					.77				-.69
.2								.63	
.25			-1.00					-.53	
.35	-1.4								
.4	-2.44		-2.19			-1.68			
.45		-1.72	-1.83		-1.76				
.5	-2.01		-3.35	-1.77	-1.36				
.55	-4.23		-2.87	-2.19	-1.41	-1.69			
.6	-3.36						-1.47		.93
.65	-3.31						-1.20		

white response probabilities correspond relatively well. Similar behaviour is exhibited in the diagnostic plots for the data sets with rates 0.6 (not shown), 0.8 (not shown), and 2.0.

The diagnostic plot for the data set with stimulus rate 1.0 indicates that for the white responses there is an overestimation by the threshold model when $U_t > \hat{\theta}_W$. The black responses are predicted quite well except for the right tail, where the empirical probabilities are slightly larger than the predicted. This may be an indication that the threshold for the black responses has been overestimated, since responses are occurring in the data with a higher than predicted probability when $\hat{U}_t > \hat{\theta}_B$. Diagnostics for the pooled data set with rate 1.2 have a similar type of behaviour (not shown). For both the white and black response probabilities the fit is quite good for positive values of $\hat{U}_t$. However, both also have a higher probability of responses occurring when $\hat{U}_t$ is smaller in magnitude than the threshold, possibly indicating overestimation of the thresholds.

The threshold model fits the data set with rate 1.4 quite well. Diagnostic plots display a good fit for the white responses, and a relatively good fit for for the black responses. The fits are also quite good for the data set with rate 4.0 (not shown), with the exception of one outlier at the left tail of the black response probability

Table 6.5: Ordinal logistic threshold estimates $\hat{\theta_B}, \hat{\theta_W}$ for choice RT data.

rate	$\hat{\theta_B}$	$\hat{\theta_W}$
0.4	-5.61	5.02
0.6	-4.92	4.50
0.8	-4.46	4.09
1.0	-4.20	3.92
1.2	-3.66	3.42
1.4	-3.53	3.43
2.0	-2.89	2.80
4.0	-2.16	2.29
8.0	-1.69	1.68

prediction. Diagnostics for the data set with rate 8.0 (not shown) are relatively good - there are a few stray empirical points that do not follow the theoretical curve that well, but the majority of the empirical points follow the theoretical curve.

Overall, the real data exhibit similar behaviour to data simulated from our parametric model; both real and simulated data show an improvement in fit as the stimulus rate increases. For slower stimulus rates it is more suggestive that our threshold model does not fit the experimental data very well. The poor fits for data with slower stimulus rates are likely because fewer non-responses occur when the flash rate is slow. In the threshold model the response probabilities are high when the magnitude of U_t is large, and U_t increases with the number of non-responses.

6.2 Marginal Models for Choice RT data

In the previous model, the incorporation of interaction effects allows inference on the presence of any inhibition or facilitation among flashes of the same type, and accounts for all of the information in the data in a single model. However, with the single impulse response, inferences cannot be made on the intensity of a specific response type following a certain kind of flash. To overcome this, we consider a marginal model for each response type. Doing so allows us to examine four impulse responses -

the "correct" impulse responses (black stimulus-black response, white stimulus-white response) and "error" responses (black stimulus-white response, white stimulus-black response).

Our conceptual model, is an extension of the one described in Section 6.1, with two internal potentials and two thresholds. Black and white flashes pass through an internal filter corresponding to the flash type, and the noisy filter outputs model the respective internal potentials. When the black internal potential exceeds the black threshold θ_B the subject decides that there is a black flash and has a 'black' response. Similarly, a 'white' response occurs when the white internal potential crosses the white threshold θ_W .

If a subject has a response bias, then the threshold corresponding to the more frequent response type is smaller than the other threshold; when $\theta_B = \theta_W$ there is no response bias.

We use the following notation:

- S^B and S^W denote the black and white stimulus point processes, respectively
- R^B and R^W denote the black and white response point processes, respectively
- X^B and X^W denote the discretized black and white stimulus point processes, respectively.
- Similarly, Y^B and Y^W denote the discretized black and white response processes, respectively.
- Occasionally $Y_t = 2$ due to the black and white buttons being pressed simultaneously. In such an event, we set $Y_t^W = 1, Y_{t+1}^B = 1$.

In order to determine the effect of black stimuli and of white stimuli on each type of response, two separate (marginal) models are fit. An internal potential is modelled for each type of response, and each type of internal potential is compared with a corresponding threshold.

For simplicity, we describe the model in terms of the white internal potential. Assuming that a white flash occurs at time τ, $a_{t-\tau}^W$ represents the effect of that flash

on the white internal potential at time t. Similarly, assuming that a black flash occurs at time τ, $b^B_{t-\tau}$ represents the effect of that flash on the white internal potential at time t. Note that time t is in terms of an index corresponding to the interval containing the time of interest. In an analogous manner we define the coefficients $a^B_{t-\tau}$ and $b^W_{t-\tau}$. As of time t, the number of time intervals elapsed since the last interval containing a response (either black or white) is denoted by γ_t. We assume that a response occurs at time 0.

We start with a linear model as an approximation to the internal potential. Under this assumption, the white and black internal potentials can be expressed as follows (Brillinger (1988b)):

$$\begin{aligned} U^W_t &= \sum_{u=0}^{g_t-1} a^W_u X^W_{t-u} + \sum_{u=0}^{i_t-1} b^B_u X^B_{t-u}, \\ U^B_t &= \sum_{u=0}^{g_t-1} a^B_u X^B_{t-u} + \sum_{u=0}^{i_t-1} b^W_u X^W_{t-u}, \end{aligned}$$

where $g_t = \min\{\gamma_t, G\}$, $i_t = \min\{\gamma_t, i\}$, and G and i are the respective indices of the corresponding interval that contains the maximum time lag of interest for flashes of the same type as the response, and of the opposite type of the response, respectively.

A white response occurs when

$$U^W_t + \varepsilon_t > \theta_W,$$

and a black response occurs when

$$U^B_t + \delta_t > \theta_B,$$

where ε_t and δ_t are internal noise, which we assume to be have a symmetric distribution about 0. More specifically, we assume a logistic error distribution.

We next consider an extension of the linear model so that interaction terms are included. As in the ordinal logit threshold models, we take into account the fact that a response to a flash may be dependent on the occurrence of previous flashes that are close together in time with the present one. That is, for the white response case we consider interactions of a white flash at time t with the occurrence of any

white flashes that are up to d^W time units earlier, for some d^W, and of a white flash occurrence at time t with the occurrence of any black flashes that are up to d^B time units earlier.

For simplicity we set $d^W = rm,\ d^B = sm$, for some integers r, s, since we are working with data that is discretized using lag m. Letting $Z_j^W = I\{\sum_{i=1}^{r} X_{j-i}^W > 0\}$ and $Z_j^B = I\{\sum_{i=1}^{s} X_{j-i}^B > 0\}$, we can express the white internal potential as follows:

$$U_t^W = \sum_{u=0}^{g_t-1} a_u^W X_{t-u}^W + \sum_{u=0}^{i_t-1} b_u^B X_{t-u}^B + \sum_{u=0}^{z_t-1} c_u^W X_{t-u}^W Z_{t-u}^W + g_u^B X_{t-u}^W Z_{t-u}^B, \qquad (6.2.1)$$

where $z_t = \min\{\gamma_t, z\}$, and $z \leq G$ is the index of the interval containing the maximum interaction (current flash occurrence with flash occurrences within d^B or d^W time units earlier) time of interest. Similarly, we can express the black internal potential as follows:

$$U_t^B = \sum_{u=0}^{g_t-1} a_u^B X_{t-u}^B + \sum_{u=0}^{i_t-1} b_u^W X_{t-u}^W + \sum_{u=0}^{z_t-1} c_u^B X_{t-u}^B Z_{t-u}^B + g_u^W X_{t-u}^B Z_{t-u}^W. \qquad (6.2.2)$$

The main coefficients have the following interpretations:

- a_u^W: impulse of white flashes and white responses
- b_u^B: impulse of black flashes and white responses
- a_u^B: impulse of black flashes and black responses
- b_u^W: impulse of white flashes and black responses,

so that a_u^W and a_u^B correspond to the "correct impulses", while b_u^B and b_u^W coincide with the "error" impulses.

Interaction effects can be used to make inferences on the existence of inhibition or facilitation among the flashes. For example, the effect on a white flash at $t-u$ by a white flash within d^W time units before it is given by c_u^W, while the effect by a black flash within d^W time units earlier is given by g_u^B. Similar interpretations hold for c_u^B and g_u^W. In all cases, a positive coefficient indicates facilitation, while a negative coefficient signifies an inhibition effect.

Each marginal model is fit by a generalized linear model of the binomial family, and the method of maximum likelihood is used to obtain estimates of the threshold and any coefficients included in the marginal model. We assume a logistic error distribution with location 0 and scale 1, so that the logit link function is employed.

We fit binomial models with logit link functions to the data in R by using the **glm** function (R Development Core Team (2006)). For the white responses, our fitted model with interactions has the form:

$$\hat{Y}_t^W = \begin{cases} 0, & \text{if } \hat{U}_t^W < \hat{\theta}_W \\ 1, & \text{if } \hat{U}_t^W > \hat{\theta}_W \end{cases},$$

where

$$\hat{U}_t^W = \sum_{u=0}^{g_t-1} \hat{a}_u^W X_{t-u}^W + \sum_{u=0}^{i_t-1} \hat{b}_u^B X_{t-u}^B + \sum_{u=0}^{z_t-1} \hat{c}_u^W X_{t-u}^W Z_{t-u}^W + \hat{g}_u^B X_{t-u}^W Z_{t-u}^B,$$

and we fit the model

$$\hat{Y}_t^W = \begin{cases} 0, & \text{if } \hat{Y}_t^{W*} < 0 \\ 1, & \text{if } \hat{Y}_t^{W*} > 0 \end{cases},$$

where

$$\hat{Y}_t^{W*} = \sum_{u=0}^{g_t-1} \hat{a}_u^W X_{t-u}^W + \sum_{u=0}^{i_t-1} \hat{b}_u^B X_{t-u}^B + \sum_{u=0}^{z_t-1} \hat{c}_u^W X_{t-u}^W Z_{t-u}^W + \hat{g}_u^B X_{t-u}^W Z_{t-u}^B + \hat{\beta}^W$$

so that the estimate of the threshold $\hat{\theta}^W$ is the negative of the intercept estimate β^W. An analogous formulation is used for the black responses.

We fit models to the same nine data sets as in the previous section, each having a different stimulus rate; the rates of flashes per second are 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 2.0, 4.0, and 8.0. We use a lag of $m = 0.05$.

We can assess goodness of fit of the model by examining plots of the empirical probability of each type of response against the corresponding predictor $\hat{U}_t^W$, $\hat{U}_t^B$. Each diagnostic plot has the same form used in assessing the fit of the threshold model for the go-no go experiment of Chapter 4.

As a means of studying the behaviour of our fitted model and the diagnostics in a situation where we know how the data is formed, we fit our threshold models to

data simulated from our parametric model for choice RT data, described in Chapter 5. The data is simulated using Algorithm 5.2.1.

For simulations run at different flash rates, the behaviour of the four impulse responses appear similar for fixed parameter settings, except for a lower peak height as the stimulus rate increases. For simulations with large values of d, there is a tendency for a small trough on the right side of the peak (located near lag μ) in the "correct" impulse responses.

The "error" impulse responses for simulations with small values of q tend to fluctuate around 0, and at slower stimulus rates a sharp positive peak usually occurs after μ. Such behaviour is observed for any d, but is more common as d increases. When the error probability q is small, flashes of the opposite type to the response cause few of that response type, which explains the fluctuation around 0. The sharp peak near μ is likely due to the identical reaction time distributions for responses to both flash types; the error responses usually occur μ time units after a flash.

As q approaches 0.5, the "error" impulse responses have a peak near μ, and exhibit behaviour similar to the "correct" impulse responses; at $q = .5$ the four impulse responses possess the same peak heights at μ. The case of $q = .5$ corresponds to a situation in which the subject responds with complete randomness to the stimuli. This also explains why for positive d troughs also appear in the "error" impulse responses. The two responses types are equiprobable, so when two flashes of the same type are presented within d time units apart, either response type is as likely to occur. Thus, the inhibition present between flashes of the same type exists for both response types, regardless of the flash type.

An example of a plot of the four stimulus-response impulses for a simulation with parameters $\mu = .5$, $\sigma = .12$, $d = .1$, $p = .1$, $p_{S_B} = p_{S_W} = 1$, $q = .05$, and $p_N = 0$, and consisting of 800 flashes, is given in Figure 6.7. In the two "correct" impulses there is evidence of a dip near a lag of 0.1 s. The two "error" impulses are near zero with the exception of a sharp peak near 0.5 for the white stimuli-black response impulse, and near 0.6 for the black stimuli-white response impulse.

Our next example is for the case of completely random responses to the flashes.

The impulses for a choice RT simulation with parameters $\mu = .5$, $\sigma = .12$, $d = .2$, $p = .1$, $pS_B = .5$, $p_{S_W} = .5$, $q = .5$, and $p_N = 0$, and consisting of 800 flashes, is given in Figure 6.8. A valley near lag 0.2 is evident in each of the plots, and the four impulses have peaks of similar height near 0.5.

Examples of threshold estimates for various simulated data sets are provided in Table 6.6. For the simulated data we expect $\hat{\theta}_B \doteq \hat{\theta}_W$ since we know that there is no response bias. The difference between the estimates appears to decrease with the flash rates, regardless of the parameter values. Furthermore, when we fix the flash rate for a simulation and allow the model parameters to vary, the threshold estimates are very similar. This observation suggests that the minimum rod current required for flash detection changes with flash rate, but is not affected by the degree of temporal summation (increases with d). As in the ordinal logistic threshold model of Section 6.1, as the flash rate increases, there is a decrease in the threshold estimates.

Table 6.6: Examples of marginal threshold estimates for simulated data. The upper value in each table element is $\hat{\theta}_B$, while the lower term is $\hat{\theta}_W$.

$d \backslash p_S$	.5	2
0	4.75	2.33
	4.48	2.36
.1	4.71	2.76
	4.49	2.72

When nonlinear inhibition is absent in the parametric model, so that the thinning parameter is equal to zero, the diagnostics indicate that for slow stimulus rates (e.g. $p_S = .5$) the fit is poor below $\hat{\theta}_B$ for the black marginal diagnostic and above $\hat{\theta}_W$ for the white diagnostic. However, as the rate increases to $p_S = 2$ the diagnostics reveal an excellent fit with the exception of a couple of outliers at the upper tail. As the thinning parameter increases to .1, the fits improve, and for faster rates such as $p_S = 2.0$ the fit is much better.

The diagnostic behaviour may be explained by the parametric model, as in the previous section; consecutive non-responses increase the response probability for the threshold model, but this is not true for the parametric model from which the data

is generated. However, as the flash rates increase the number of non-responses in a time period also increase, so that there appears to be a positive relationship with the response probabilities. As d increases this relationship appears more apparent, which is also explained by the fact that in our parametric model, the number of non-responses increases with d. Examples of diagnostic plots with fast and slow flash rates are given in Figures 6.9 and 6.10, respectively.

We now discuss the fits of our marginal threshold models to the experimental data. For almost all data sets we study, the effects of interactions between opposite types of stimuli are not different from zero in the "best-fitting" model for each internal potential. Therefore, such interactions are only included in the data set with rate 1.2, which is the only one for which the fit was improved by their inclusion.

Plots of the impulse responses of the "black" button responses to black stimuli and "white" button responses to white stimuli, as well as the error responses "black"-white and "white"-black are provided in Figures 6.11 and 6.13, respectively. Plots of the impulse response peaks and error response peaks as a function of the lag are also displayed in Figures 6.12 and 6.14. There is a tendency for the peaks of the "black"-black and "white"-white impulses to decrease with the stimulus rate. This is consistent with intuition. With a slower stimulus rate one is less likely to miss a stimulus or mistakenly press the wrong button. A higher peak is an indication of a larger frequency of the response type to the stimulus type at the lag for which the peak occurs. The error impulses have smaller peak values; with a lower frequency the wrong button is pressed.

The fitted marginal models of the internal potentials U_t^B, U_t^W for the nine pooled data sets follow, where a lag of $m = .05$ was used. For illustration purposes, we include terms that are negligibly different from zero in Figures 6.11 and 6.13, but in the fitted models given below we only include those that are significantly different from zero. In the output from R a p-value is provided for each coefficient. We set the level of significance as $\alpha = .05$, so that those with a p-value smaller than 0.05 are included in our model. The coefficients of the black and white stimulus effects included in the models are given in Tables 6.7 and 6.8.

Table 6.7: Black stimulus coefficients included in black marginal threshold models fit to choice RT data.

Lag \ rate	0.4	0.6	0.8	1.0	1.2	1.4	2.0	4.0	8.0
0					-.86				
.05									
.1									.77
.15									
.25					.65	.74			
.3	1.23	.79	.86		1.40	1.41	.62		
.35	2.26	1.59	2.19	2.14	.90	1.52	.82		
.4	3.28	2.32	2.49	1.90	1.31	1.60	1.43		
.45	3.26	2.19	2.69	2.21	2.26	1.94	2.20		
.5	2.26	3.37	3.70	3.14	2.04	2.46	2.00		
.55	4.70	3.80	3.66	3.04	2.28	2.80	2.05		
.6	4.69	3.94	3.44	3.50	1.97	2.45	2.58		
.65	4.67	3.29	3.52	3.03	1.78	1.87	2.08		
.7	4.38	3.68	3.51	2.22	1.29		3.30		
.75	4.48	2.68	3.64	3.31	1.41				
.8	5.05	3.11	2.37	3.16	2.23				
.85	4.43	3.83	4.28						
.9	4.78	2.89	3.69						
.95	3.86								
1.00			3.80						

The black flash interaction coefficients included in the black threshold model are given in Table 6.9. There tends to be facilitation among the black flashes at lags smaller than 0.3, and inhibition at larger lags. This inhibition is not evident in data sets with flash rates higher than 1.4, and is most common in the data set with rate 0.8. In the rate 0.8 data set the inhibition effect appears to be at lags centered around the mean reaction time, 0.5.

Table 6.10 gives the white flash interaction coefficients that are included in the white marginal model. As in the case of the black flashes, there tends to be facilitation among the white flashes at lags smaller than 0.3, and inhibition at larger lags. This inhibition is not evident in data sets with flash rates higher than 1.2, and is most common in the data set with rate 0.4. In the rate 0.4 data set the inhibition effect appears to be at lags before the mean reaction time, 0.5.

Table 6.8: White stimulus coefficients included in white marginal threshold models fit to choice RT data.

Lag \ rate	0.4	0.6	0.8	1.0	1.2	1.4	2.0	4.0	8.0
0					-.86				
.05									
.1							-1.10		.54
.15						-1.23			
.25					.65	.74			
.3		1.16	.97	.75	1.25	.89	1.22		
.35	.97	1.84	1.90	1.64		1.06	.67		
.4	2.61	2.45	1.97	2.14	1.26	1.81	.95		
.45	2.77	2.66	2.20	2.04	1.99	2.51	1.74		
.5	2.84	3.76	2.96	3.41	2.10	2.61	2.15	2.73	
.55	4.16	3.75	2.96	3.72	2.79	2.34	1.25		
.6	4.16	3.77	3.45	3.22	2.45	3.00	1.82		
.65	4.37	3.47	2.17	3.02	2.92	1.54	2.64		
.7	4.13	3.84	3.70	2.70	1.79	1.83	2.76		
.75	3.02	3.95	2.82		1.46				
.8	3.37	3.70	2.68	1.91	2.90				
.85	3.24	2.63		2.92	3.38				
.9	1.49	4.08	3.30						
.95	1.85								
1.00	2.70	3.88	3.80						
1.05	3.37		1.05						

Table 6.11 displays the two threshold estimates, $\hat{\theta}_W, \hat{\theta}_B$ for models that fit the data reasonably well, based on the diagnostic plots. For a given data set the two threshold estimates are quite close, so that in general we have $\hat{\theta}_W \doteq \hat{\theta}_B$. Such results suggest that there is no bias in response choice. As observed for the simulated data, the thresholds tend to decrease with the stimulus rate.

Diagnostic plots are provided for four of the data sets in Figure 6.15, as an example. For the data set with the slowest rate 0.4, the diagnostic indicates a good fit for the black responses, but suggests a poor one for white responses. For $\hat{U}_t^W > \hat{\theta}_W$, the threshold model overestimates the probability of a white response. This overestimation behaviour is also observed in both the black and white diagnostics for the data sets with rate 0.6 (not shown) and rate 0.8 (not shown), although not as poor

Table 6.9: Black flash interactions included in black marginal threshold models for choice RT data.

Lag \ rate	0.4	0.6	0.8	1.0	1.2	1.4	2.0	4.0	8.0
0							1.18	-1.18	
.1				1.51				.80	
.15						1.16			
.2		2.02			1.27			1.11	1.82
.25		2.47							
.3	-2.67								
.35			-2.37	-2.09					
.4			-2.47				-2.42		
.45			-3.89		-1.95				
.5		-3.82	-2.45						
.55			-3.79						
.6			-3.20						
.65									
.7									
.75									
.8									

for the rate 0.8 diagnostic. There is still an overestimation tendency in both diagnostics for the rate 1.0 data set, but the fit is better than the slower rate data sets. The fit is very good for white responses in the 1.2 rate data set, but there is still an overestimation problem with the black threshold fitted model. The fitted models for the data sets with rates faster than 1.2 are all excellent, according to the diagnostics. In these four cases, the empirical probabilities follow the theoretical curve quite well. This improvement in fit agrees with the results of our simulations.

Table 6.10: White flash interactions included in white marginal threshold models for choice RT data.

Lag \ rate	0.4	0.6	0.8	1.0	1.2	1.4	2.0	4.0	8.0
.15	2.63					2.32		1.23	
.2						1.43	1.46	1.60	
.3								1.90	
.35	-1.51								
.4	-2.42		-2.48	-2.02					
.45	-2.89		-2.31		-2.74				
.5	-3.74			-3.42					

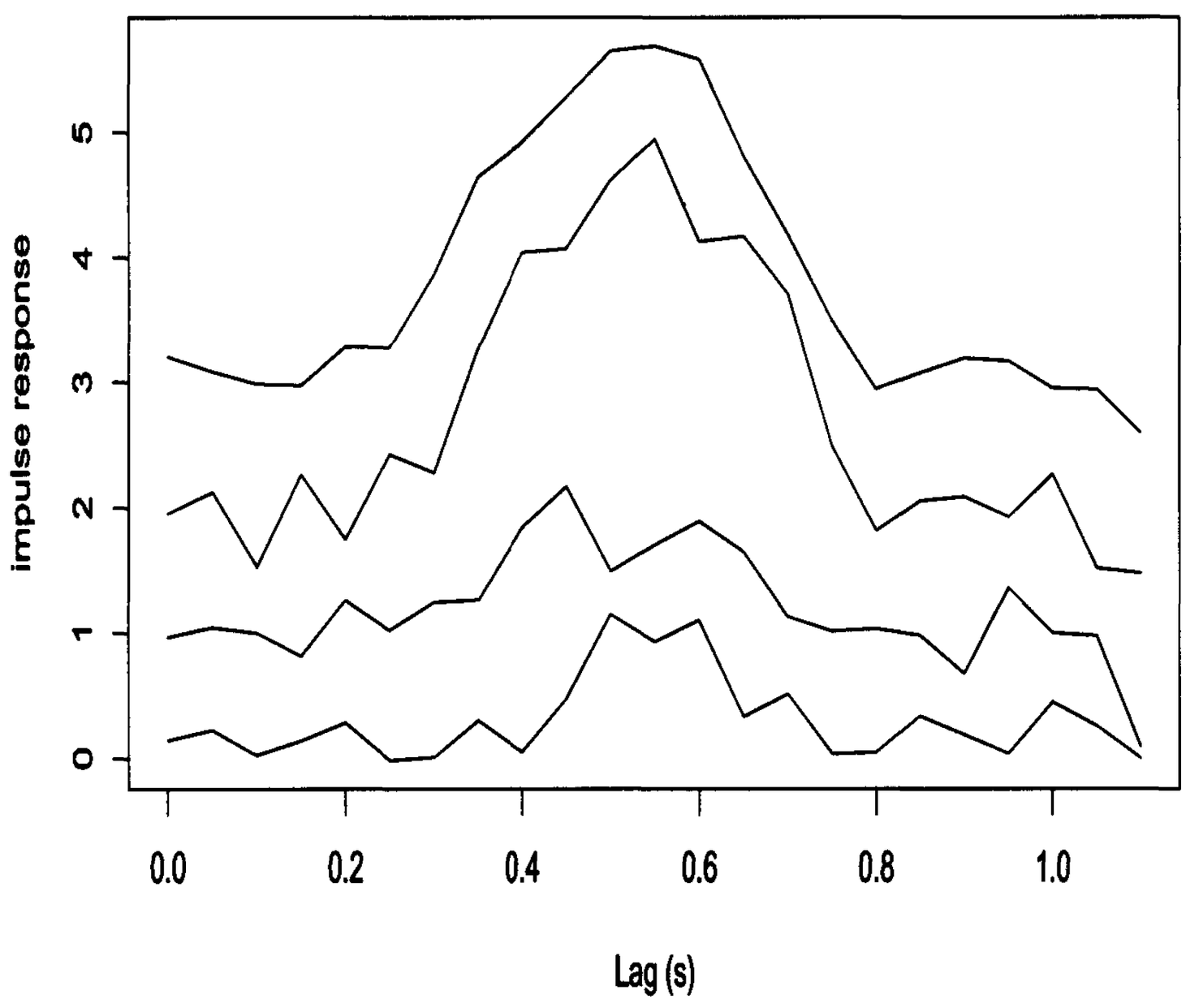


Figure 6.1: Ordinal logistic threshold model linear filters for 4 choice RT simulated data sets consisting of 800 flashes. For display purposes, each curve is shifted upwards by one unit. Starting from bottom to top, the curves correspond to simulated data sets with $(p_S = 2, d = .2)$, $(p_S = 2, d = 0)$, $(p_S = .5, d = .1)$, and $(p_S = .5, d = 0)$, respectively. The other parameters are held constant as $\mu = .5$, $\sigma = .12$,$p = .1$, and $q = .05$.

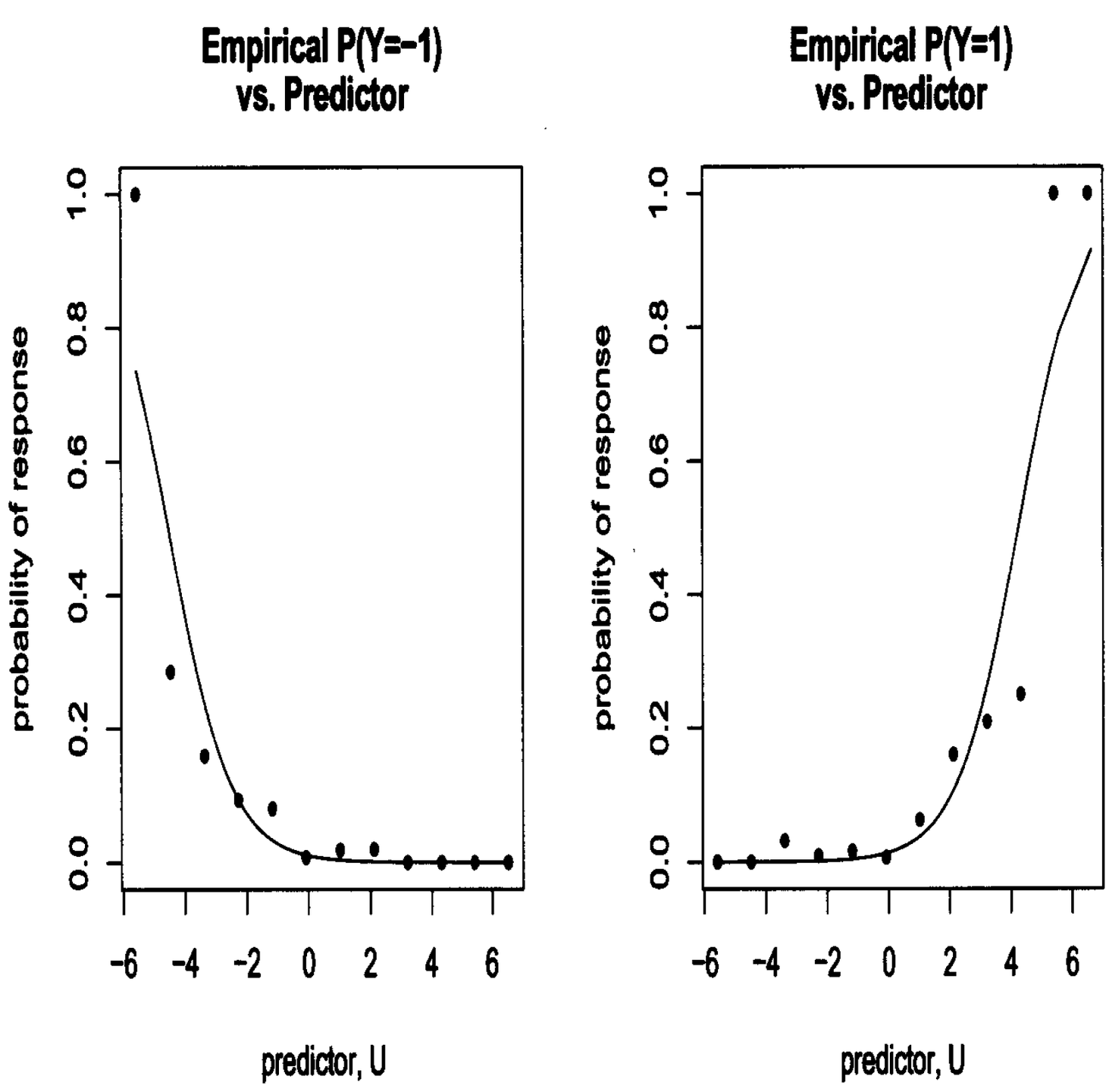


Figure 6.2: Diagnostic plots for a choice RT simulation from our parametric model with $p_S = .5$, $d = 0$, $\mu = .5$, $\sigma = .12$, $p = .1$, and $q = .05$.

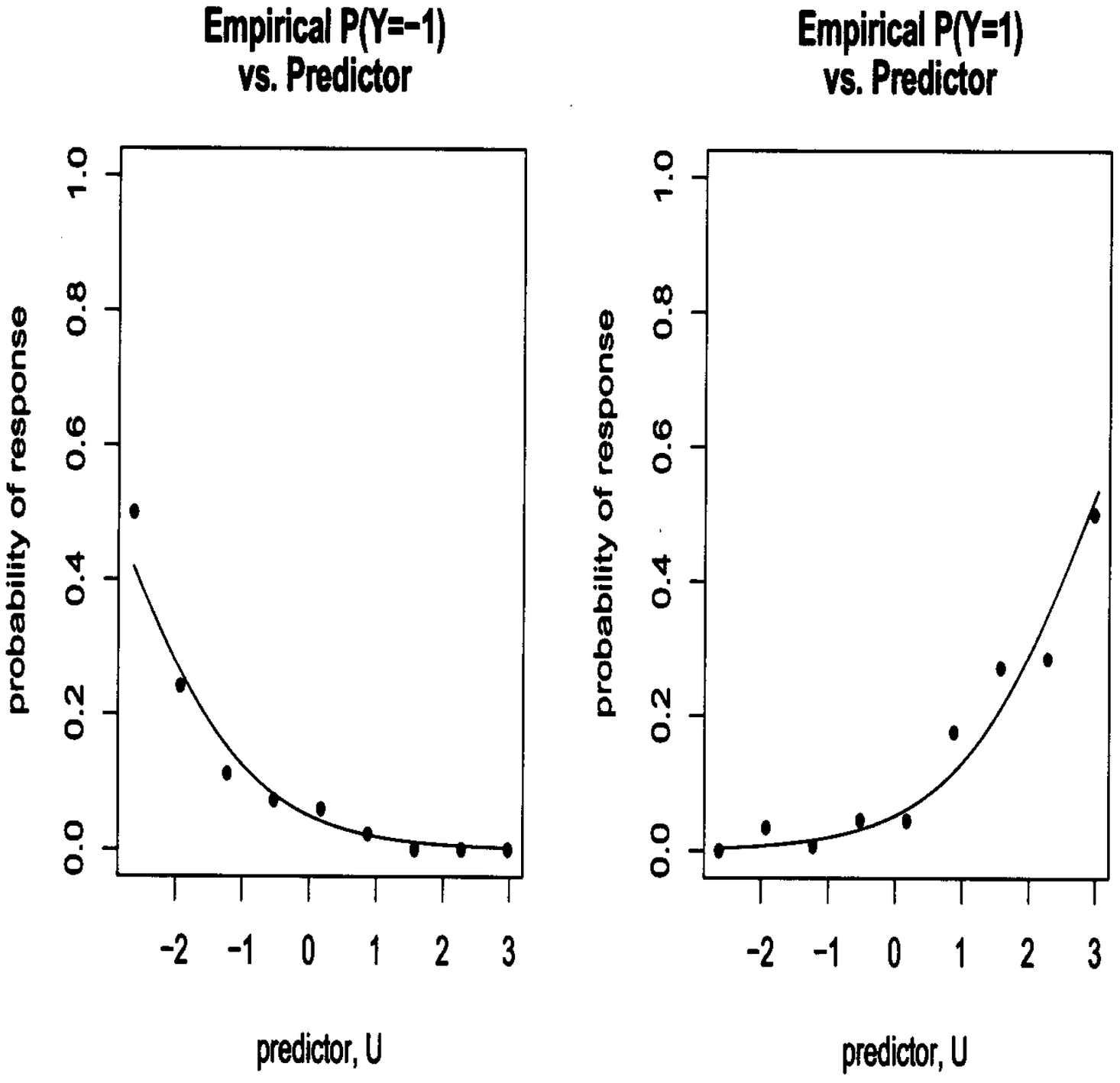


Figure 6.3: Diagnostic plots for a choice RT simulation from our parametric model with $p_S = 2$, $d = .2$, $\mu = .5$, $\sigma = .12$, $p = .1$, and $q = .05$.

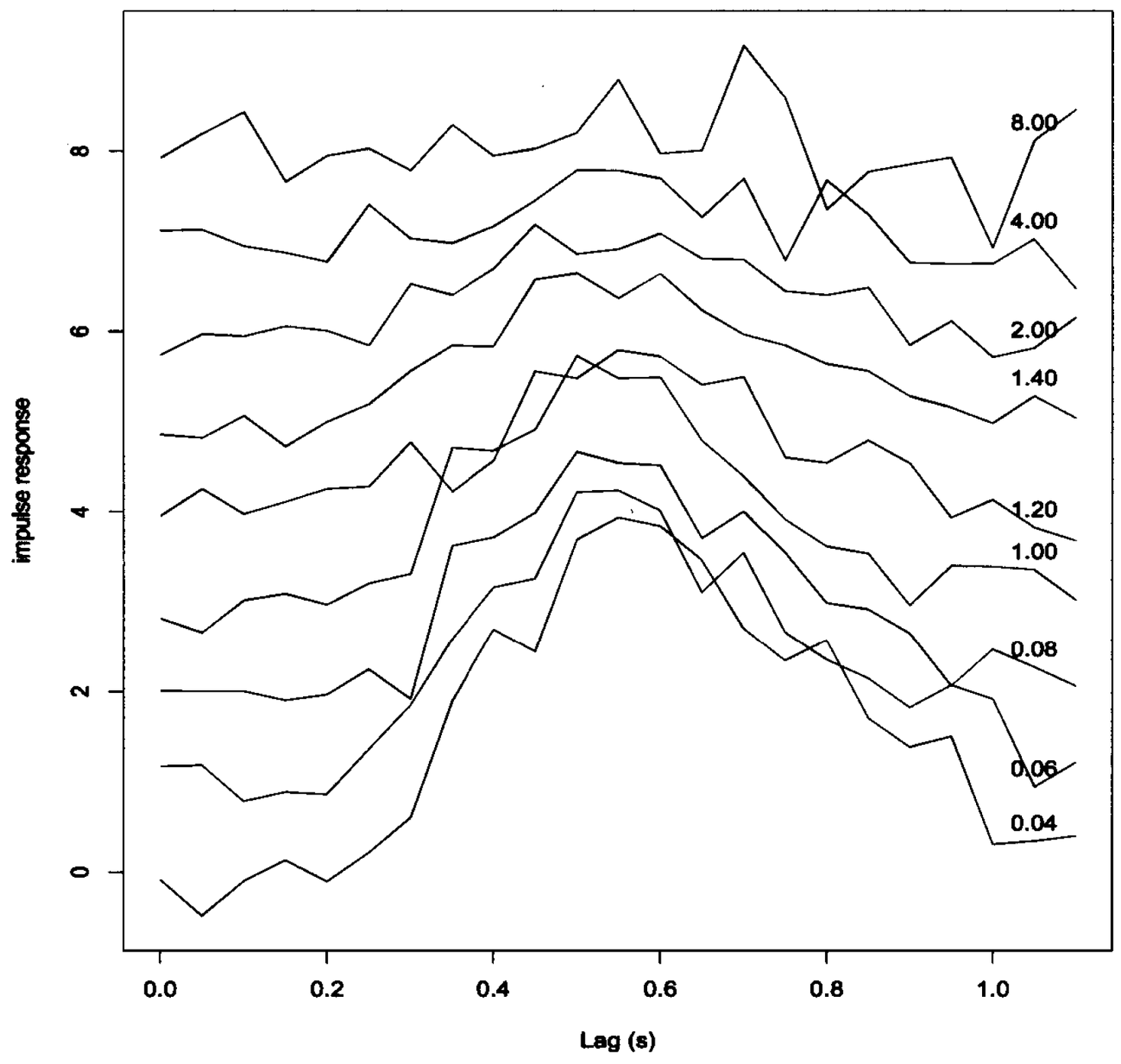


Figure 6.4: Ordinal logistic threshold model linear filters for each of the 9 choice RT data sets. Curves are labelled by the stimulus rate (s). For display purposes, each curve is shifted upwards by one unit.

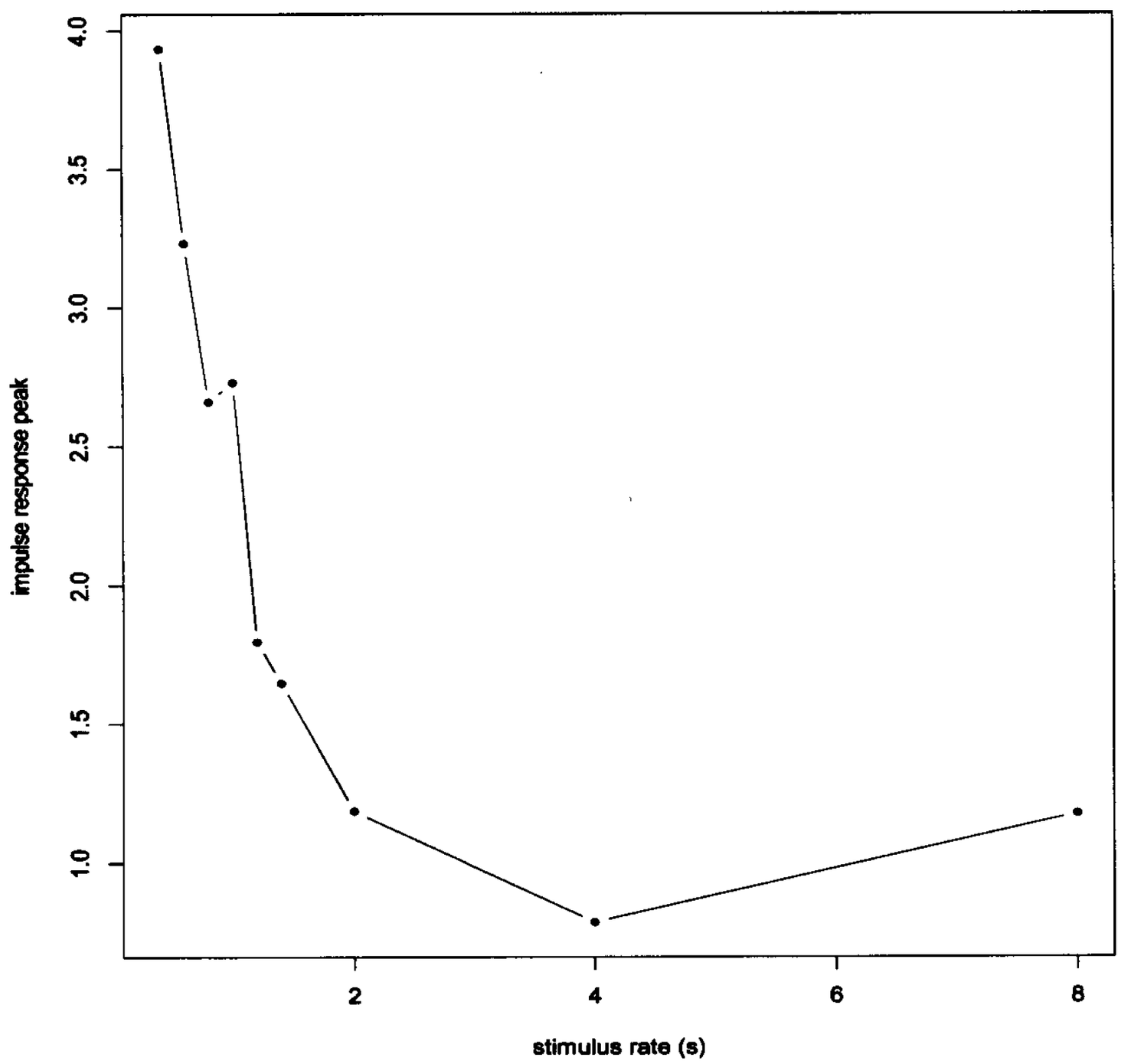


Figure 6.5: Plot of the peak height of the linear filters as a function of stimulus rate, for ordinal logistic threshold models fit to experimental data.

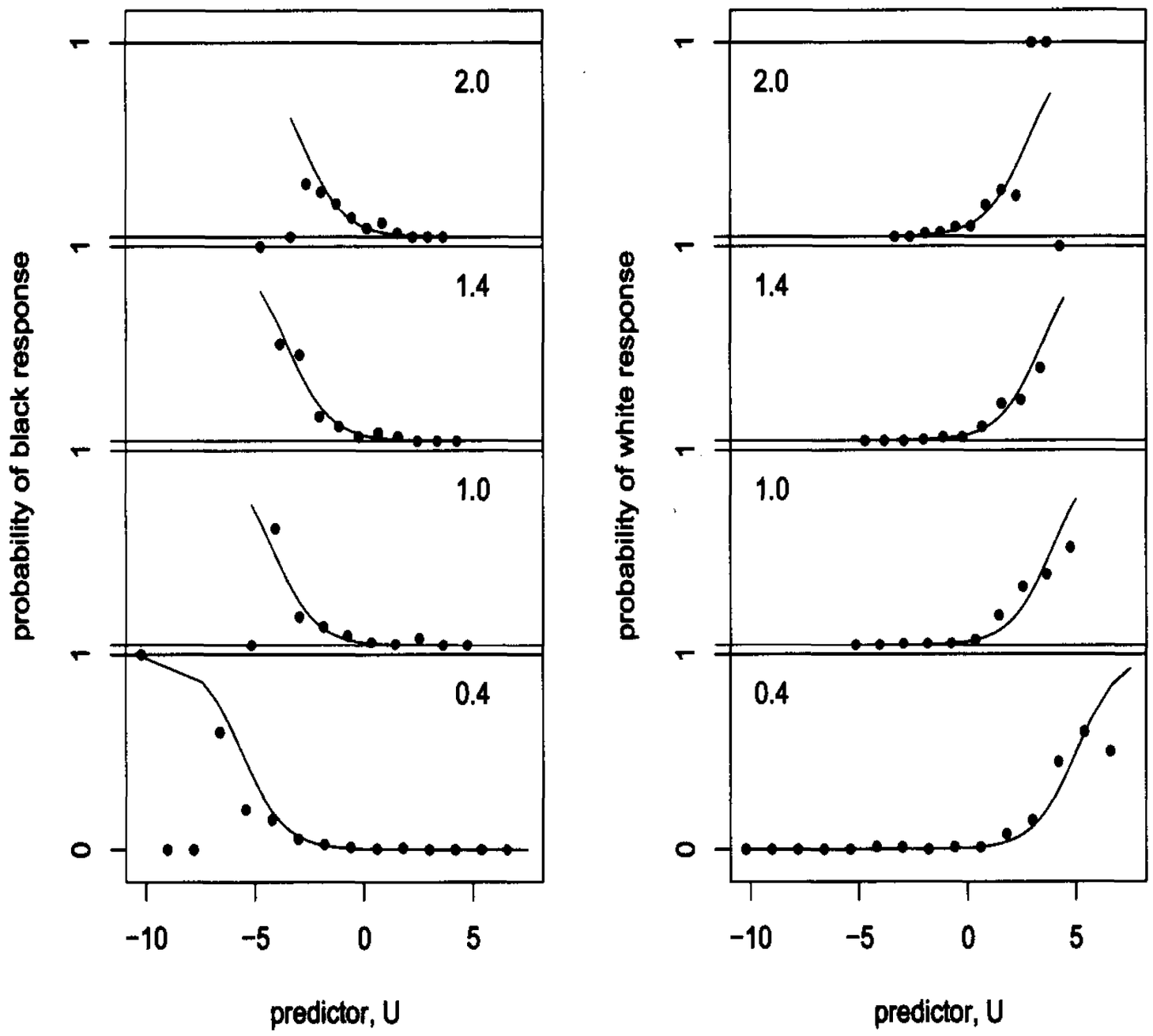


Figure 6.6: Examples of ordinal logistic threshold model diagnostic plots for the data sets with stimulus rates 0.4,1.0,1.4 and 2.0, as indicated in the top left corner of each plot. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

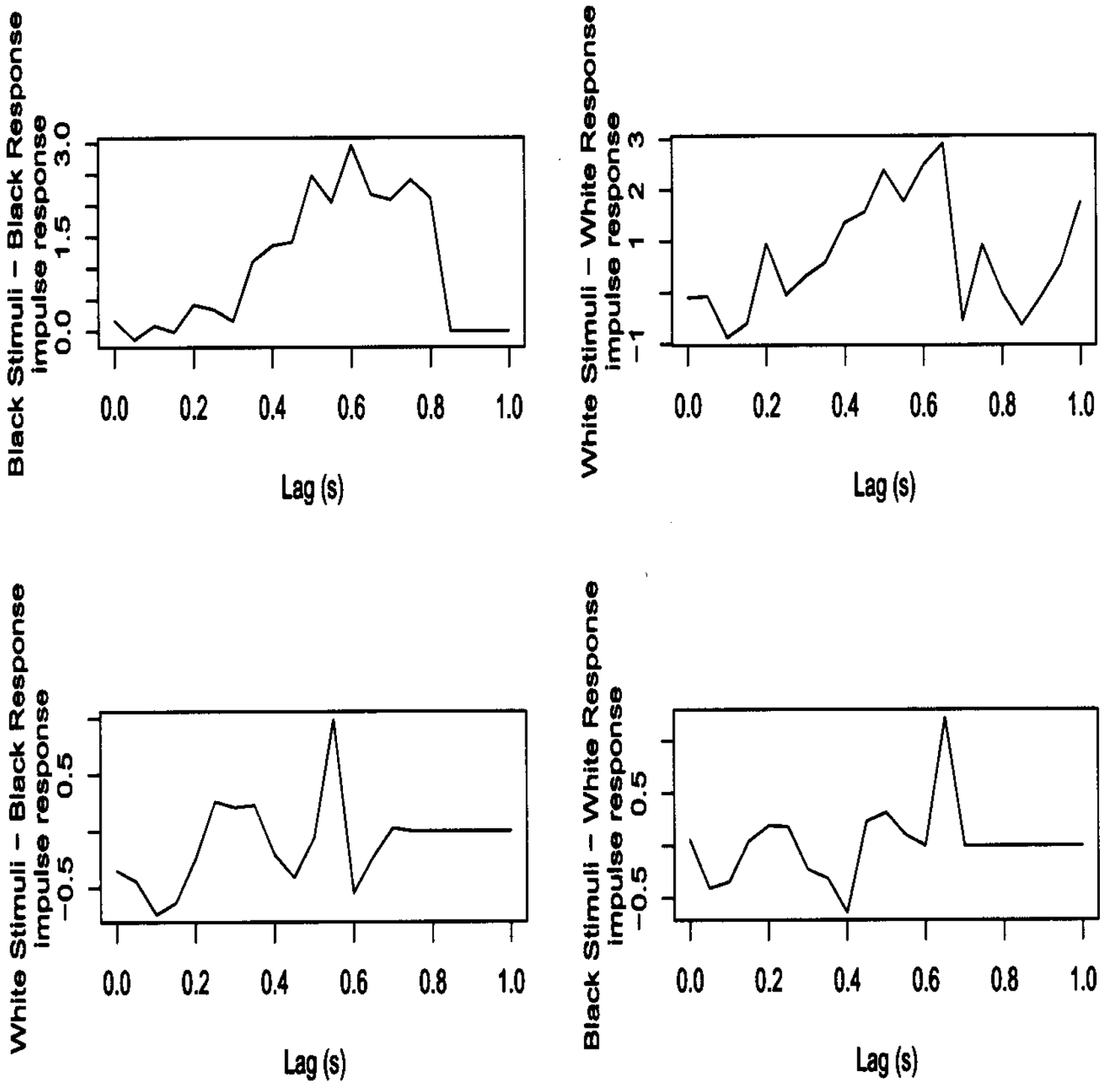


Figure 6.7: Marginal threshold model stimulus-response impulse plots for a choice RT simulation from our parametric model with $p_S = 1$, $\mu = .5$, $\sigma = .12$, $d = .1$, $p = .1$, $p_{S_B} = p_{S_W} = 2$, and $q = .05$.

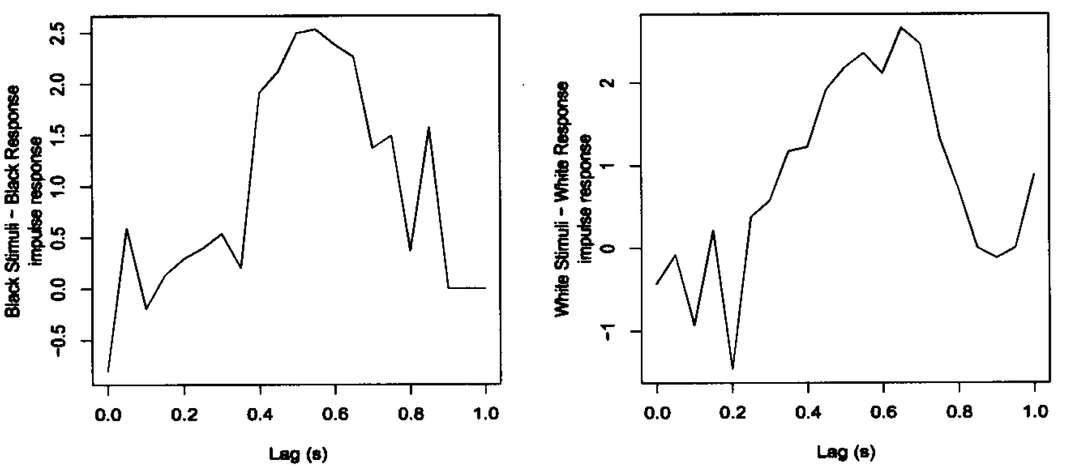


Figure 6.8: Stimulus-Response impulse plots for a choice RT simulation from our parametric model with $\mu = .5$, $\sigma = .12$ $p_S = .5$, $p = .1$, $d = .1$, and $q = .5$.

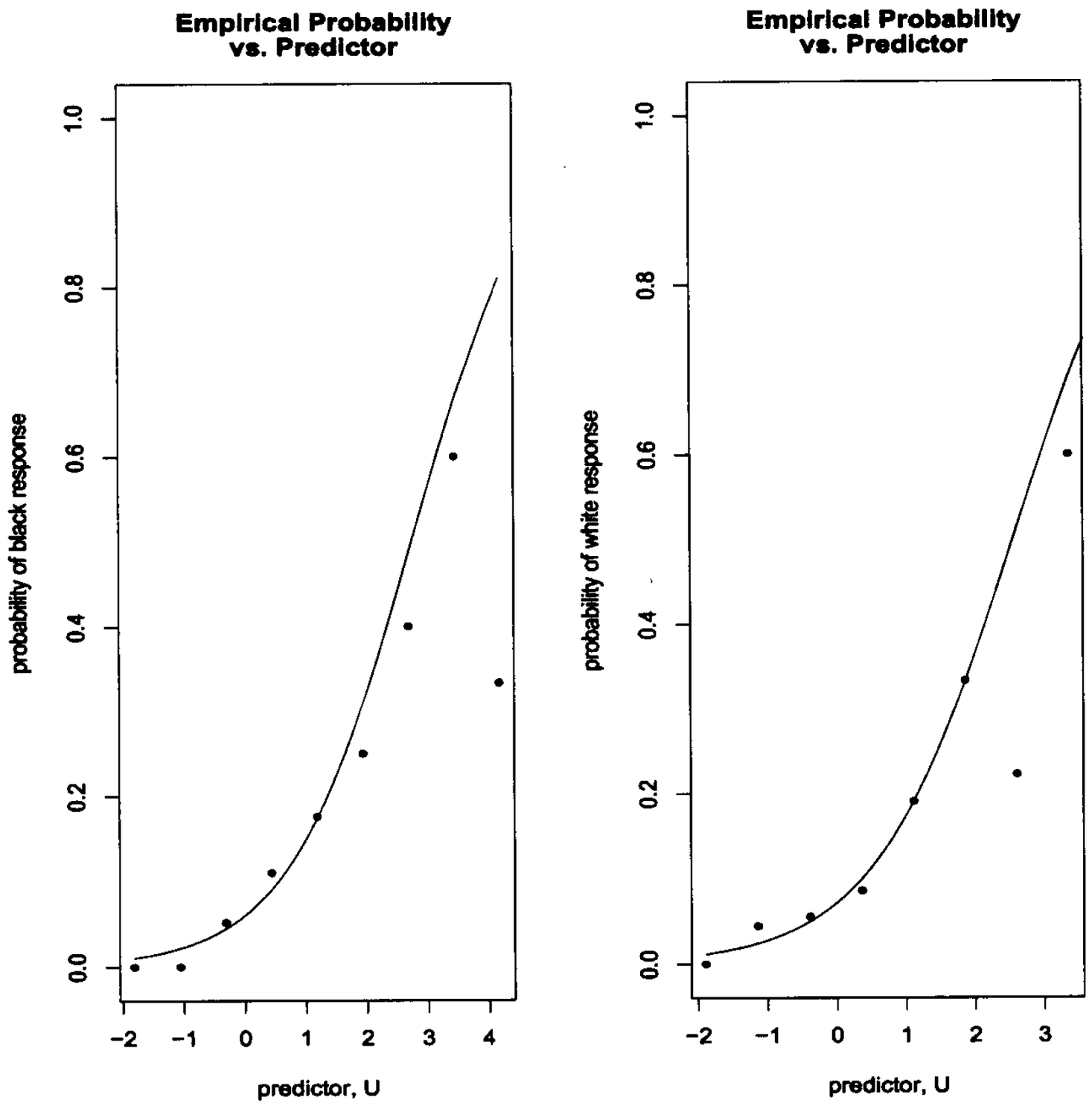


Figure 6.9: Diagnostic plots for marginal threshold models fit to a choice RT simulation from our parametric model with $p_S = 2$, $\mu = .5$, $\sigma = .12$, $d = .1$, $p = .1$, $p_{S_B} = p_{S_W} = 2$, and $q = .05$.

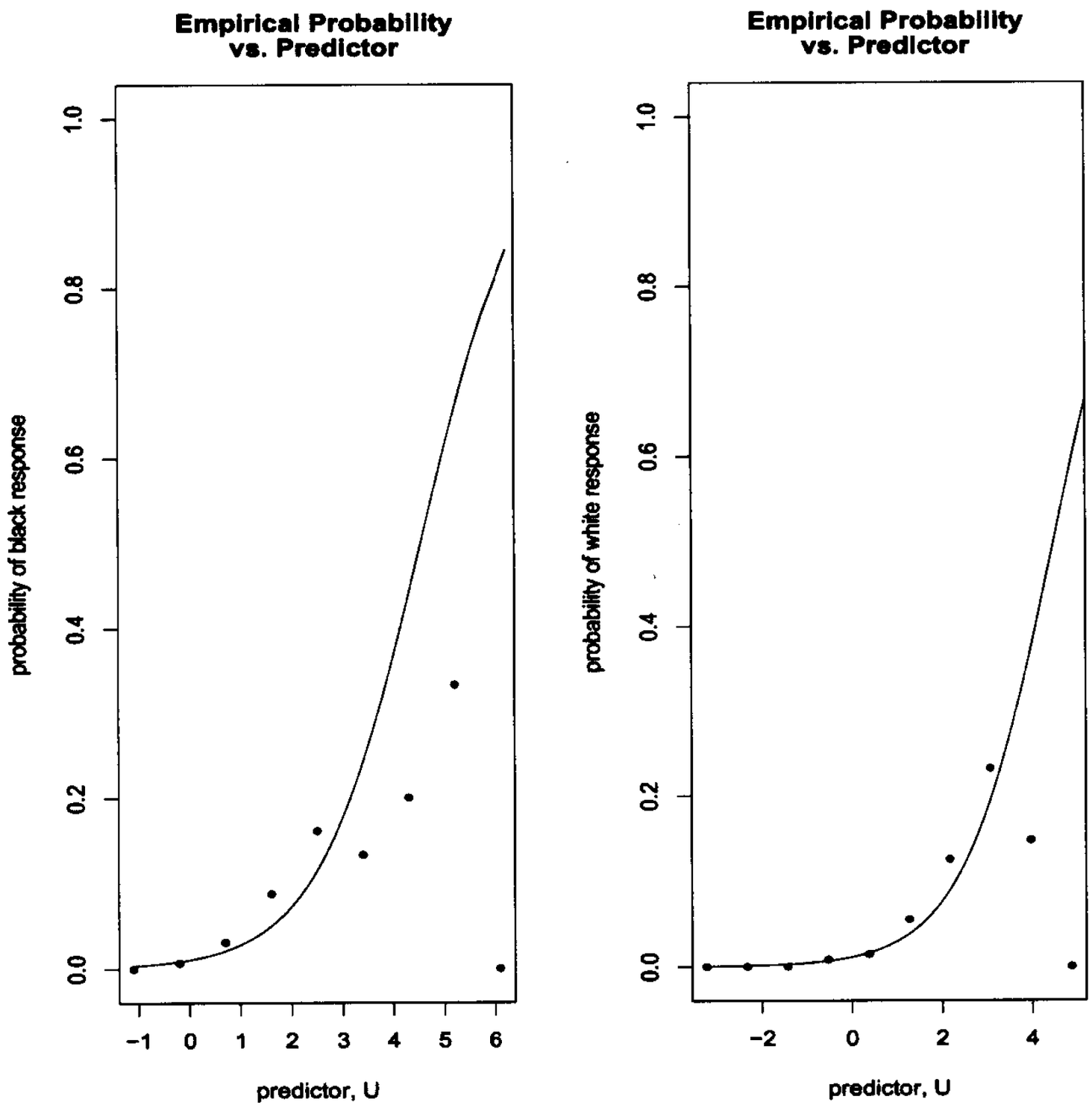
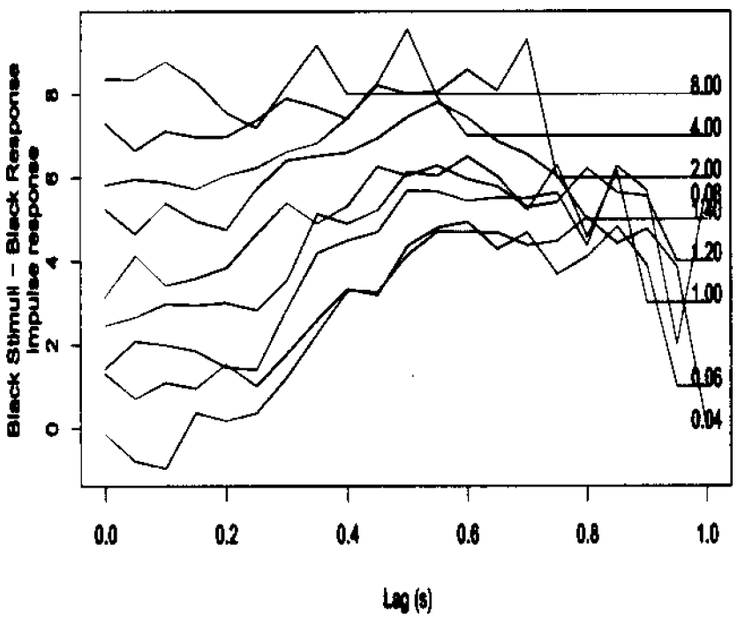
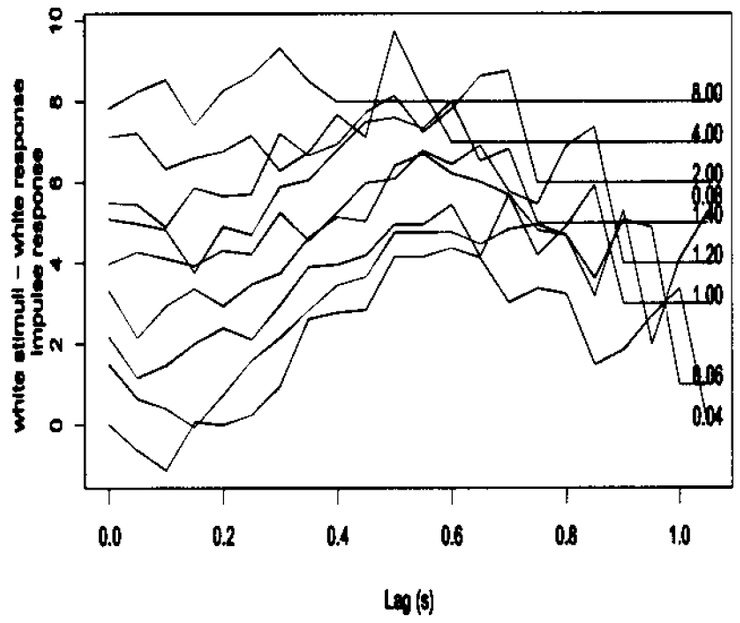


Figure 6.10: Diagnostic plots for marginal threshold models fit to a choice RT simulation from our parametric model with $\mu = .5$, $\sigma = .12$ $p_{S_B} = p_{S_W} = .5$, $p = .1$, $d = .2$, and $q = .5$.

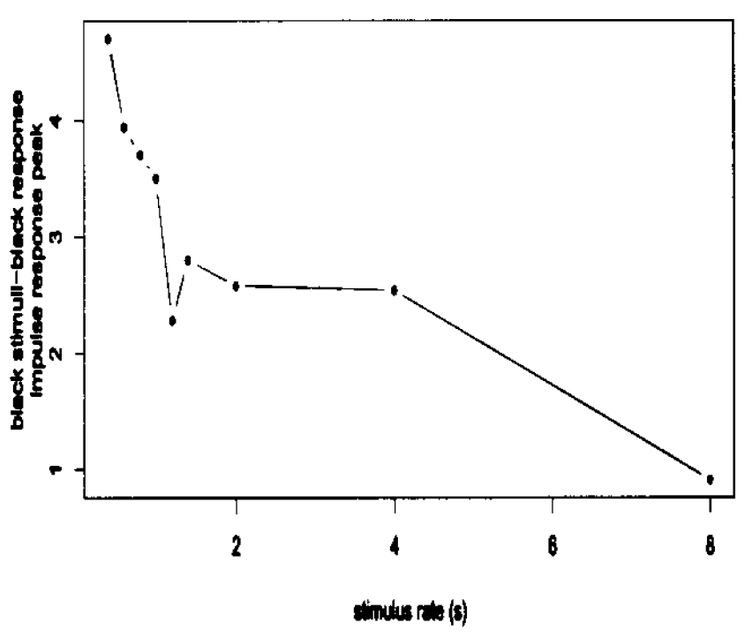


(a) Black stimulus-black response linear filter

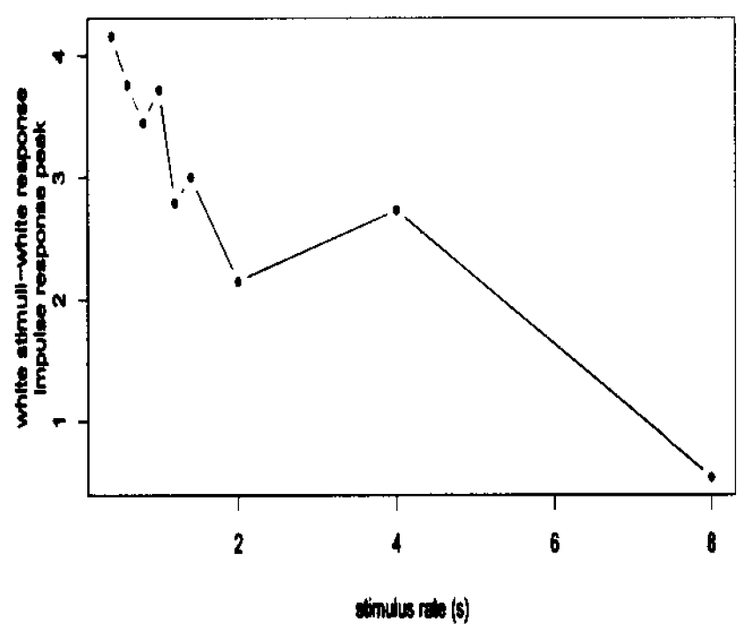


(b) White stimulus-white response linear filter

Figure 6.11: Plots of the marginal threshold "correct" linear filters for each of the 9 choice RT data sets. Curves are labelled by the stimulus rate (s). For display purposes, each curve is shifted upwards by one unit.

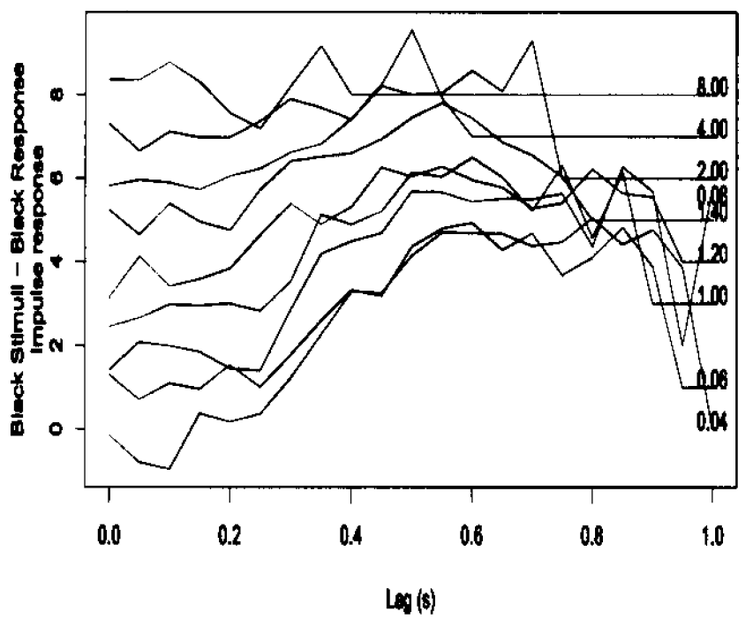


(a) Black stimulus-black response linear filter peaks

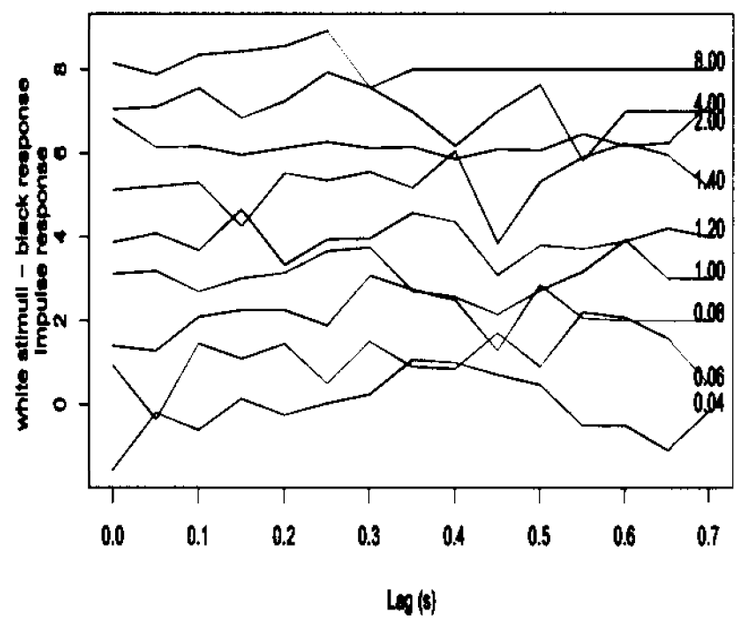


(b) White stimulus-white response linear filter peaks

Figure 6.12: Plots of the marginal threshold "correct" impulse peak heights as a function of stimulus rate, for the nine data sets.

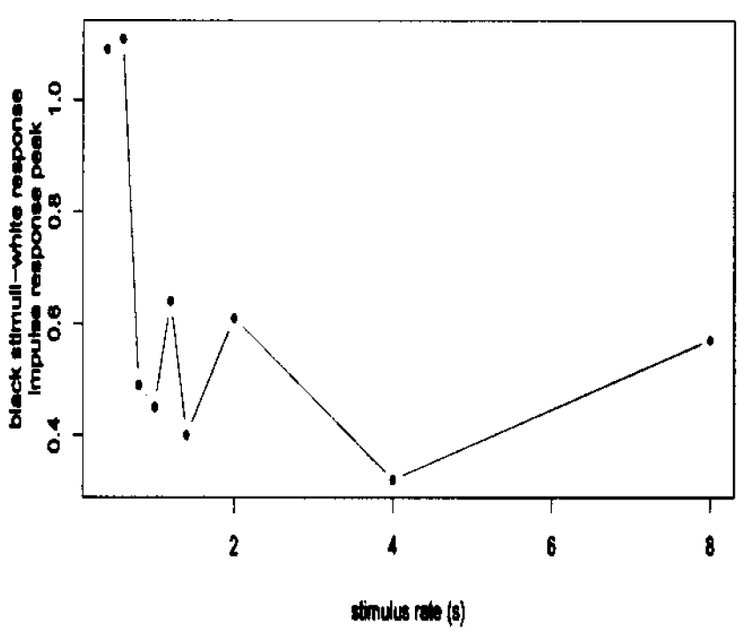


(a) Black stimulus-white response linear filter

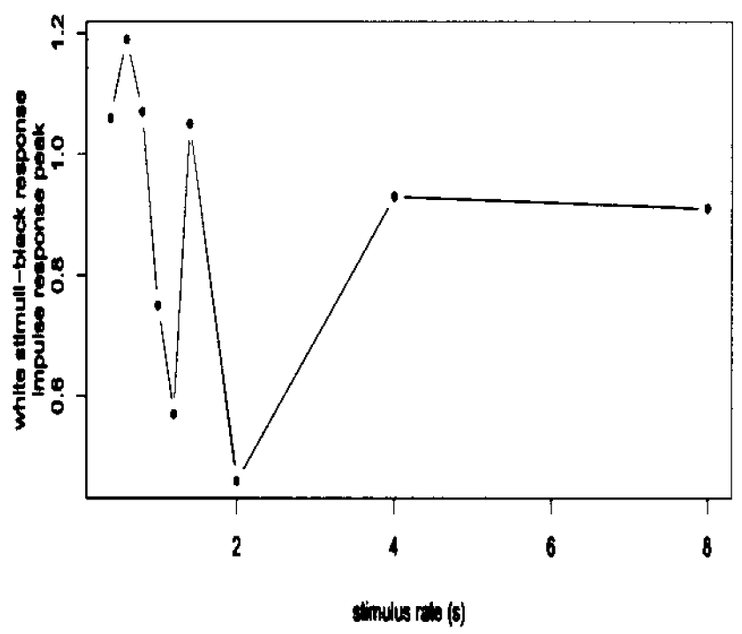


(b) White stimulus-black response linear filter

Figure 6.13: Plots of the marginal threshold "error" linear filters for each of the 9 choice RT data sets. Curves are labeled by the stimulus rate (s). For display purposes, each curve is shifted upwards by one unit.



(a) Black stimulus-white response linear filter peaks



(b) White stimulus-black response linear filter peaks

Figure 6.14: Plots of the marginal threshold "error" impulse peak heights as a function of stimulus rate, for the nine data sets.

Table 6.11: Threshold estimates $\hat{\theta_B}, \hat{\theta_W}$ for choice RT data.

rate	$\hat{\theta}_B$	$\hat{\theta}_W$
0.4	5.58	5.26
0.6	4.78	4.82
0.8	4.49	4.31
1.0	4.06	4.10
1.2	3.63	3.64
1.4	3.58	3.61
2.0	2.96	2.81
4.0	2.25	2.26
8.0	1.84	1.73

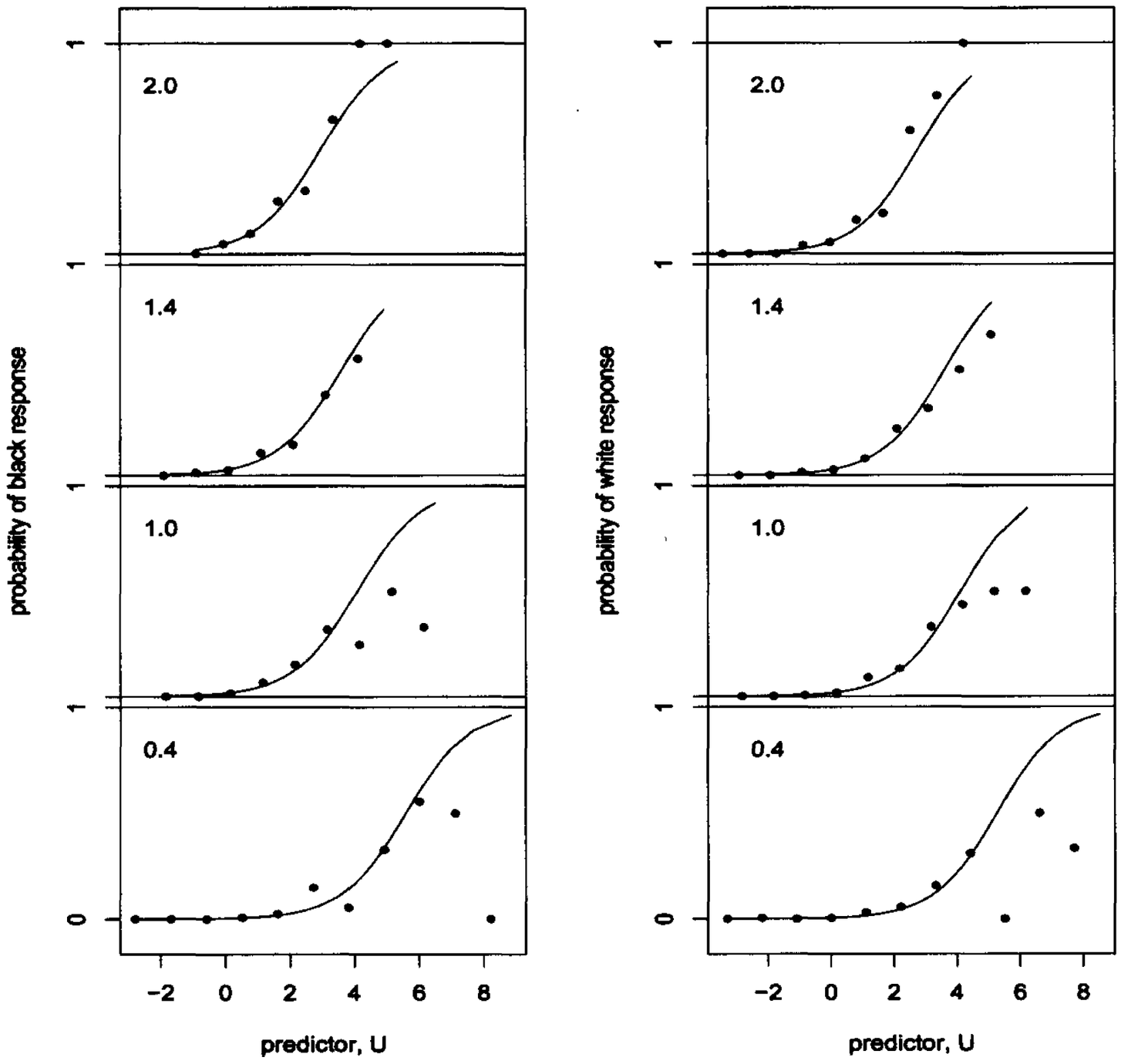


Figure 6.15: Example of marginal threshold diagnostic plots for the data sets with stimulus rates 0.4,1.0,1.4 and 2.0, as indicated in the top left corner of each plot. The points are the empirical probability of a response, and the curve is the corresponding fitted probability.

6.2.1 A Marginal Model with Second Order Interaction Effects

Interaction terms are included in the marginal model that we have studied, but only between a current flash and an indicator of whether or not a flash occurred some amount of time units before it. Now, we consider an extension of the marginal models to include interaction terms between all combinations of flashes. Doing so allows us to study second order intensities.

In the case of white responses, the internal potential can be expressed as follows:

$$\begin{aligned} U_t^W &= \sum_{u=0}^{g_t-1}(a_u X_{t-u}^W + b_u X_{t-u}^B) + \sum_{u=0}^{g_t-1}\sum_{v=u+1}^{g_t-1}(c_{u,v}X_{t-u}^W X_{t-v}^W + d_{u,v}X_{t-u}^B X_{t-v}^B) \\ &+ \sum_{u=0}^{g_t-1}\sum_{v=u+1}^{g_t-1}(r_{u,v}X_{t-u}^W X_{t-v}^B + s_{u,v}X_{t-u}^B X_{t-v}^W), \end{aligned} \tag{6.2.3}$$

where the interaction coefficient interpretations are:

- $c_{u,v}$: intensity with which a white response occurs u time units after a white flash and v time units after another white flash
- $d_{u,v}$: intensity with which a white response occurs u time units after a black flash and v time units after another black flash
- $r_{u,v}$: intensity with which a white response occurs u time units after a white flash and v time units after a black flash
- $s_{u,v}$: intensity with which a white response occurs u time units after a black flash and v time units after a white flash.

By symmetry, it follows that $c_{u,v} = c_{v,u}$, $d_{u,v} = d_{v,u}$, $r_{u,v} = s_{v,u}$, and $s_{u,v} = r_{v,u}$. The diagonal elements are undefined and we set them equal to zero. The black internal potential has an expression similar to (6.2.3) with analogous interpretations for its interaction coefficients.

In the same manner as the simpler marginal models studied we fit binomial models with logit link functions to the data in R by using the `glm` function (R Development

Core Team (2006)). However, due to the four bivariate coefficients considered a large amount of computing memory is required in R. To solve such memory problems, rather than fitting the full model including all four interaction combinations of the two flash types, we fit two separate models before the reduced model fitting. For each response type, a model is fit with white-white and black-black interactions, and one with white-black and black-white interactions. The coefficients with p-values smaller than 0.25 are included in a reduced model involving all four types of interactions. In the two reduced models (one for black responses and one for white responses) with all four second-order interaction effects, bivariate coefficients with large p-values and estimates near 0 are left in the model instead of fitting another reduced model.

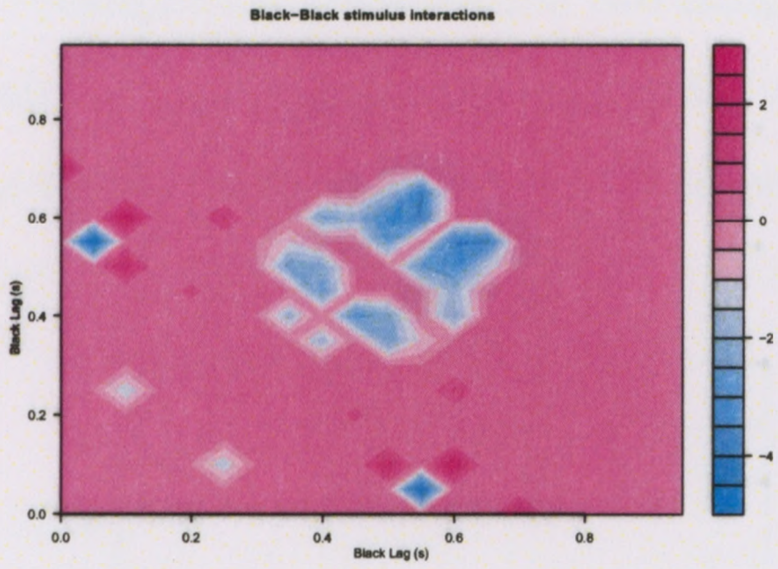
The model is assessed using diagnostics plots in the exact same way as the simpler marginal models studied earlier.

Earlier, we studied the behaviour of the fitted marginal threshold model without second-order interaction terms, and its diagnostics in depth. Therefore, due to the memory problems and computational time, we study one example of the marginal threshold model with second-order interactions fit to one simulation of choice RT data. We use Algorithm 5.3.1 to generate the data.

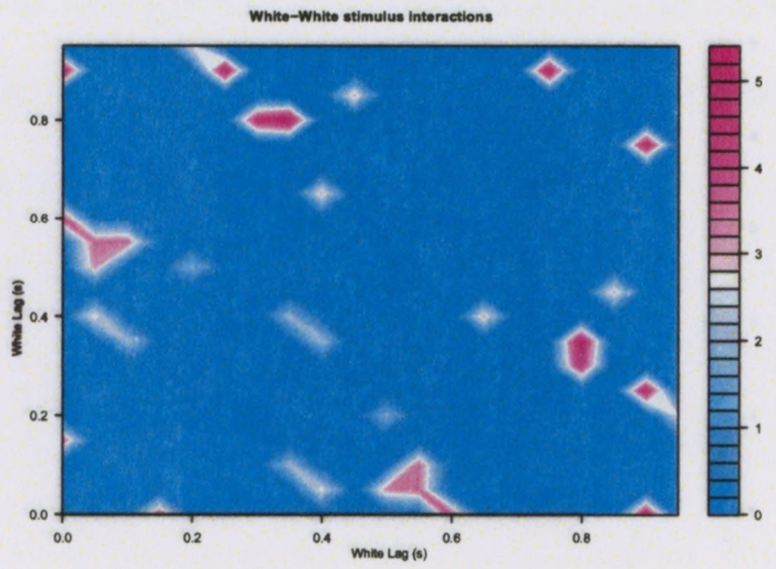
We fit models to simulated choice RT data with parameters $p_{S_B} = p_{S_W} = .5$, $\mu = .5$, $\sigma = .1$, $p = .12$, $d = .2$, $q = .1$, $p_N = 0$, and consisting of 800 flashes. Filled contour plots of the interaction effects for black responses are provided in Figures 6.16 and 6.17, while plots for the white responses are given in Figures 6.18 and 6.19.

Based on Figure 6.16(a) there appears to be frequent inhibition among black flashes when the two flashes occur between .3 s and .6 s before a black response. The data was simulated with a thinning parameter of $d = .2$, so when two black flashes are within .2 s apart the first of the two flashes is not responded to. In addition, the mean reaction time is 0.5 s. Therefore, it is reasonable that inhibition is frequent when two black flashes are within 0.2 s apart and occur approximately 0.5 s before a response. There is also evidence of facilitation when one black flash occurs 0.6 s before a black response and another one occurs less than 0.1 s before the response.

In Figure 6.16(b) there is a positive interaction effect among white flashes occur-

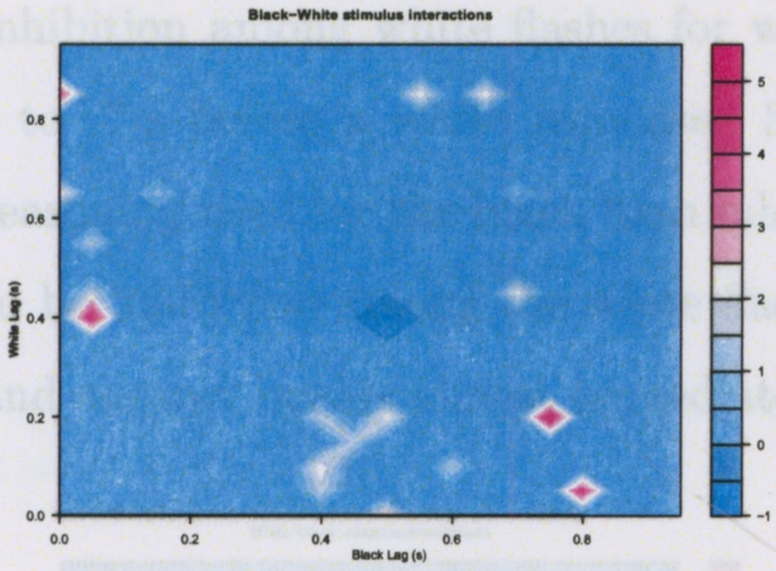


(a) Black-Black flash interactions

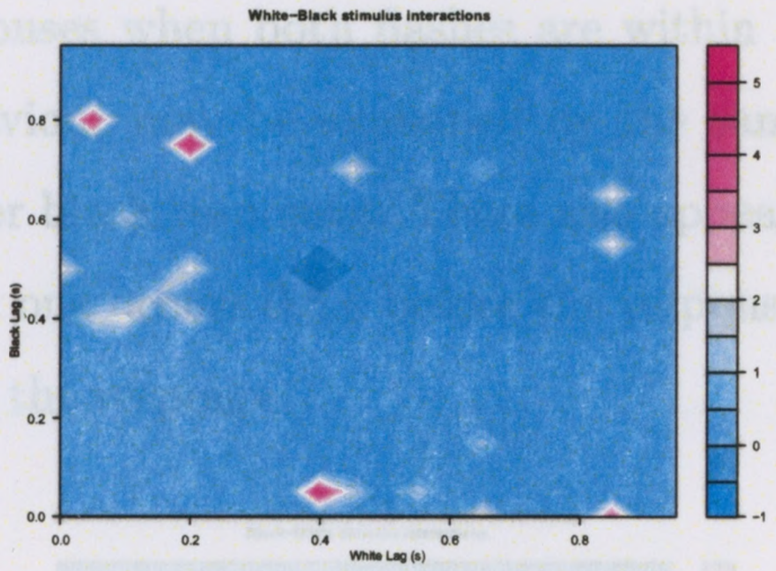


(b) White-White flash interactions

Figure 6.16: The second order intensities of the fitted black marginal threshold model for one set of choice RT simulated data with rate $p_S = .5$, $d = .2$ and $q = .1$.



(a) Black-White flash interactions



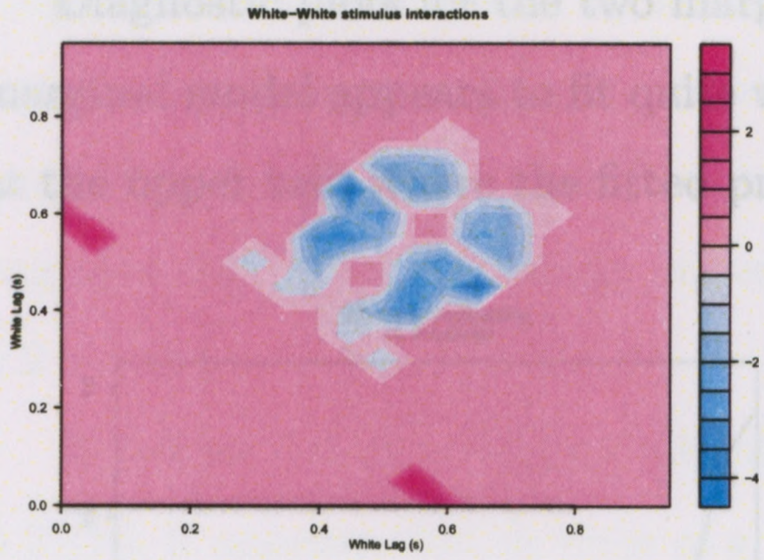
(b) White-Black flash interactions

Figure 6.17: Second order intensities of the fitted black marginal threshold model for one set of choice RT simulated data with rate $p_S = .5$, $d = .2$ and $q = .1$.

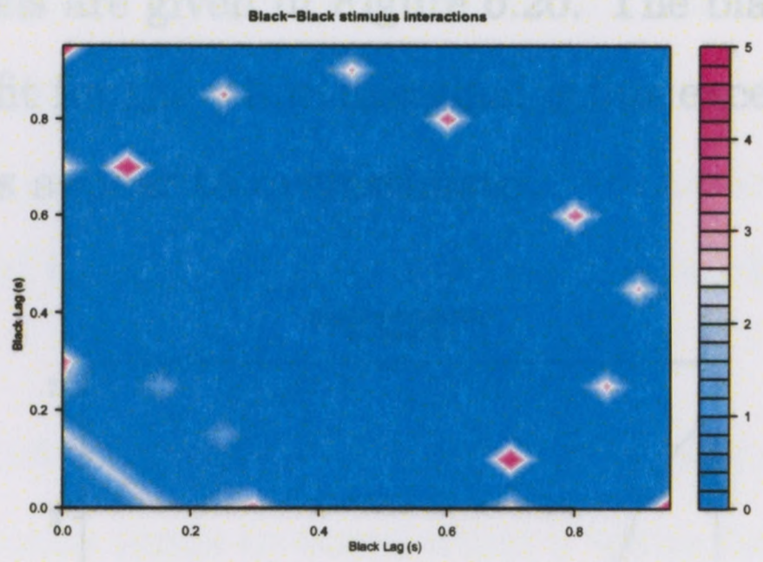
ring before a black response. This tends to occur when both flashes are around 0.5 s before a black response, and are likely due to the error probability of $q = .1$ for our simulated data; in our simulated data with probability $q = .1$ a black response occurs from a white flash, and the mean reaction time is 0.5.

In the two mixed flash interaction plots for black responses, given in Figure 6.17, there appears to be a negative interaction effect on the black responses when one of each flash type occurs around 0.5 s before a response. This may be linked to the fact that the mean reaction time from both black and white flashes is 0.5 s.

Figure 6.18(a) displays similar behaviour to 6.16(a). There is evidence of frequent



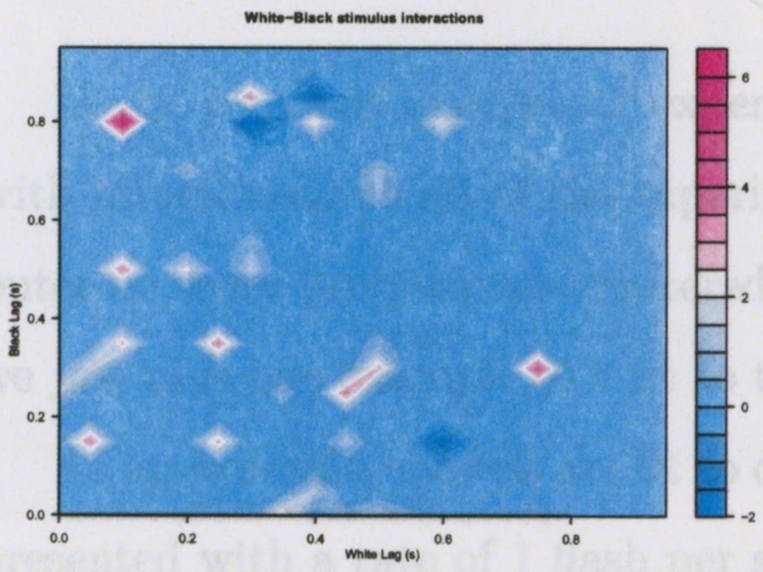
(a) White-White flash interactions



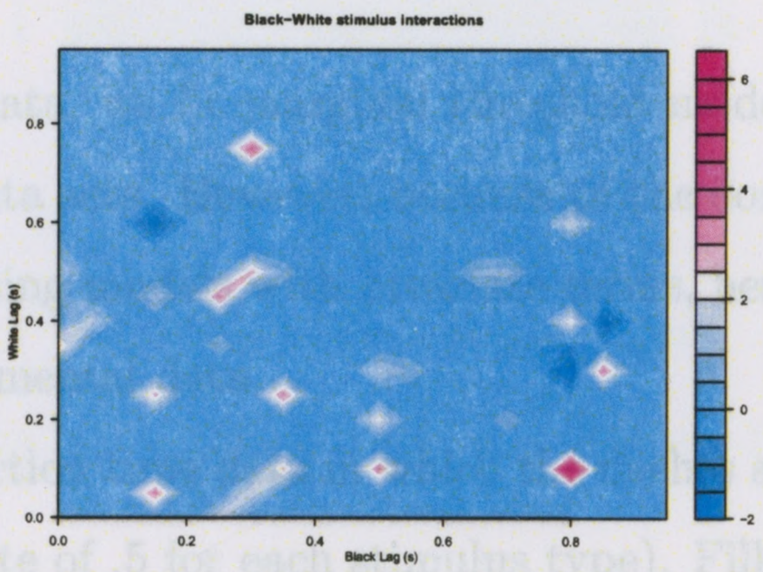
(b) Black-Black flash interactions

Figure 6.18: The second order intensities of the fitted white marginal threshold model for one set of choice RT simulated data with rate $p_S = .5$, $d = .2$ and $q = .1$.

inhibition among white flashes for white responses when both flashes are within .4 s to .7 s before a white response. Such behaviour may be explained by the same reasoning used for the black flash inhibitions for black responses. There also appears to be inhibition among the white flashes when one occurs 0.5 s before the response, and another occurs almost immediately before the response.



(a) White-Black flash interactions

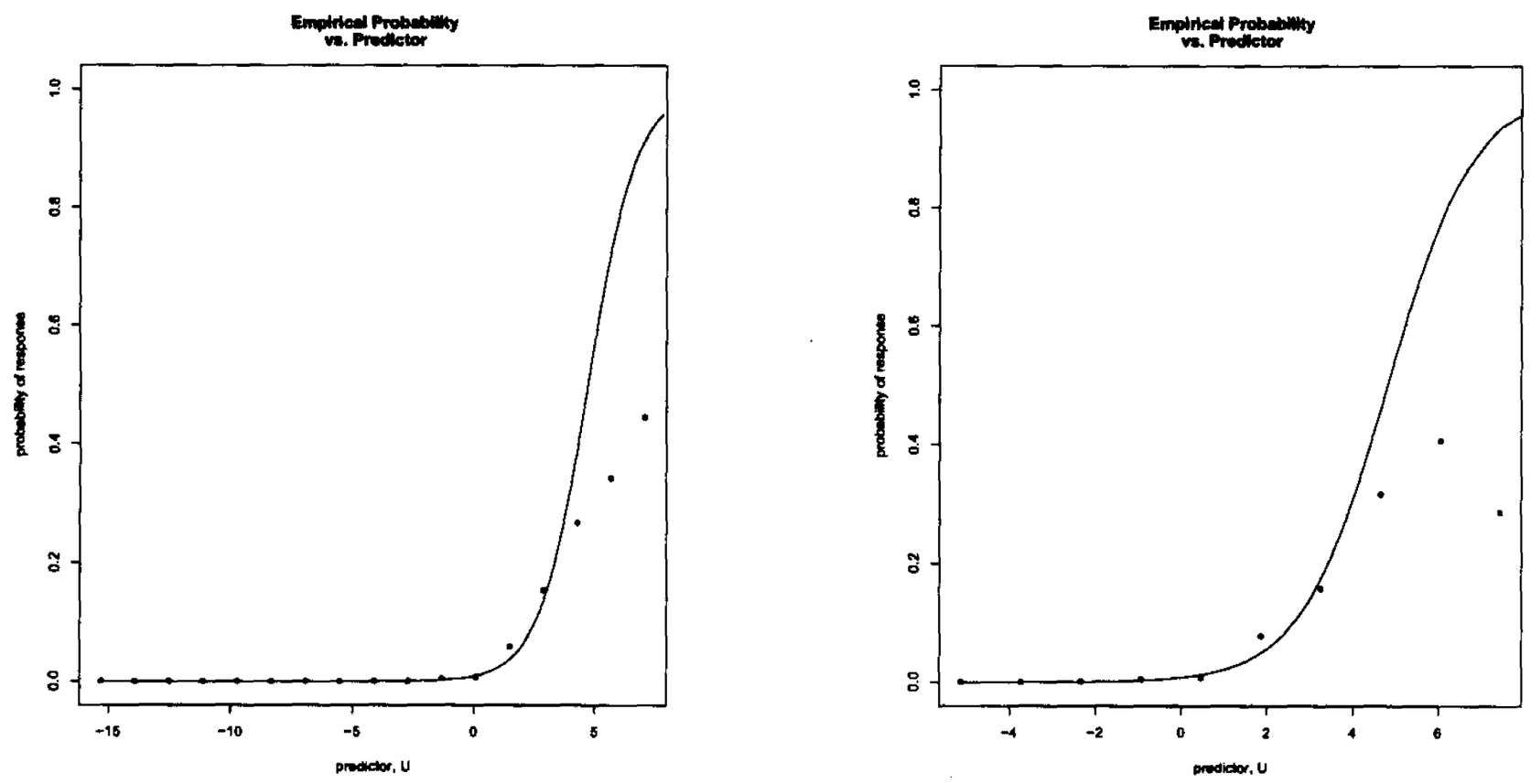


(b) Black-White flash interactions

Figure 6.19: The second order intensities of the fitted white marginal threshold model for one set of choice RT simulated data with rate $p_S = .5$, $d = .2$ and $q = .1$.

There appears to be facilitation among the black flashes for white responses when one black flash occurs around 0.7 s before a white response, and another black flash occurs nearly 0.1 s before the response. In the mixed flash interaction plots of Figure 6.19 similar behaviour is exhibited as in the case of black responses.

Diagnostic plots for the two marginal models are given in Figure 6.20. The black marginal model appears to fit quite well. The fit for the white marginal is fine except at the upper tail, where the fitted probabilities appear to overestimate.



(a) Black Response Marginal Model

(b) White Response Marginal Model

Figure 6.20: Diagnostic plots for one set of choice RT simulated data with rate $p_S = .5$, $d = .2$ and $q = .1$.

In the previous analysis of experimental data, we fit marginal threshold models with interactions to all of the experimental data sets. However, because of the computer memory problems encountered when fitting models with bivariate terms, here, we just give one example of a fit to the experimental data.

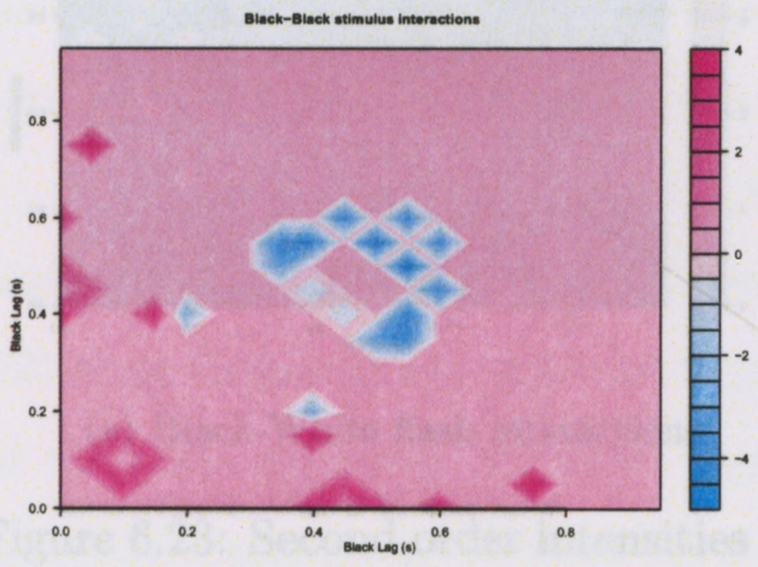
As an example, models are fit to choice reaction time data in which the flashes are presented with a rate of 1 flash per second (rate of .5 for each stimulus type). Filled contour plots of the interaction effects for black responses are provided in Figures 6.21 and 6.23, while plots for the white responses are given in Figures 6.22 and 6.24.

The threshold estimates are $\hat{\theta}_W = 4.16$ and $\hat{\theta}_B = 4.05$, which are similar enough to suggest that there is no response bias. For both the white responses and for the black responses the plots suggest that inhibition effects are present for flashes of the same type as the response. For example, in the case of the black internal potential many of the interaction effects of black flashes with each other are large and negative

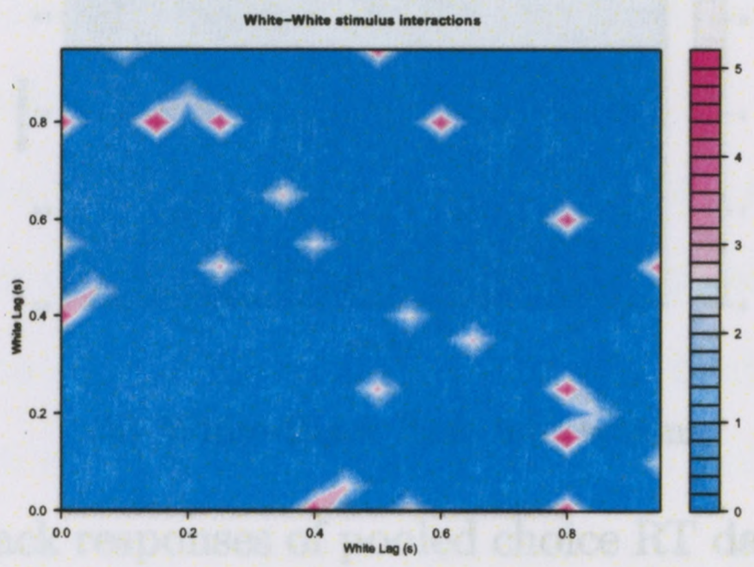
when flashes are separated by less than 0.2 seconds and occur between 0.4 and 0.6 seconds before a response. A similar pattern appears in the case of white responses and white flash interactions. Since the mean reaction time is 0.5, such behaviour suggests that there is thinning among flashes of the same type when they are within 0.2 s apart.

Based on the white-white flash interaction contour plot for black responses, Figure 6.21(b), it appears that an error of a black response tends to occur when a white flash occurs 0.8 s earlier and 0.2 s before the current time. This same error seems to occur when a white flash occurs 0.8 s earlier and 0.6 s before the current time.

Errors of a white response tend to occur when two black flashes were presented 0.65 and 0.2 seconds earlier or when two black flashes were presented around 0.35 s and 0.45 s, as can be seen from Figure 6.22(b).



(a) Black-Black flash interactions

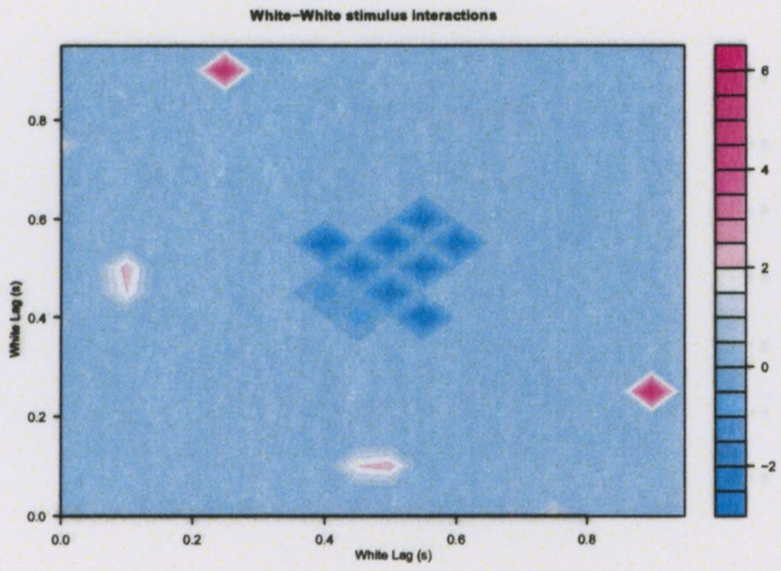


(b) White-White flash interactions

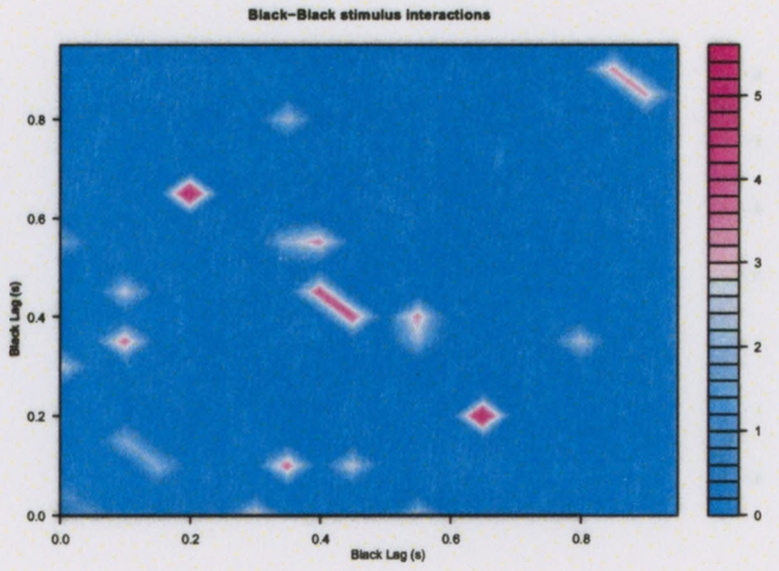
Figure 6.21: The second order intensities for the black responses of pooled choice RT data with rate $p_S = .5$.

For either response type, there does not appear to be any negative interactions between flashes of different types. There are a few instances of positive interactions for both response types, with a slightly higher frequency for white responses.

Diagnostic plots for the two marginal models are given in Figure 6.25. The white marginal model appears to fit relatively well, while the black marginal model does not fit well at the upper tail, where the fitted probabilities appear to overestimate.

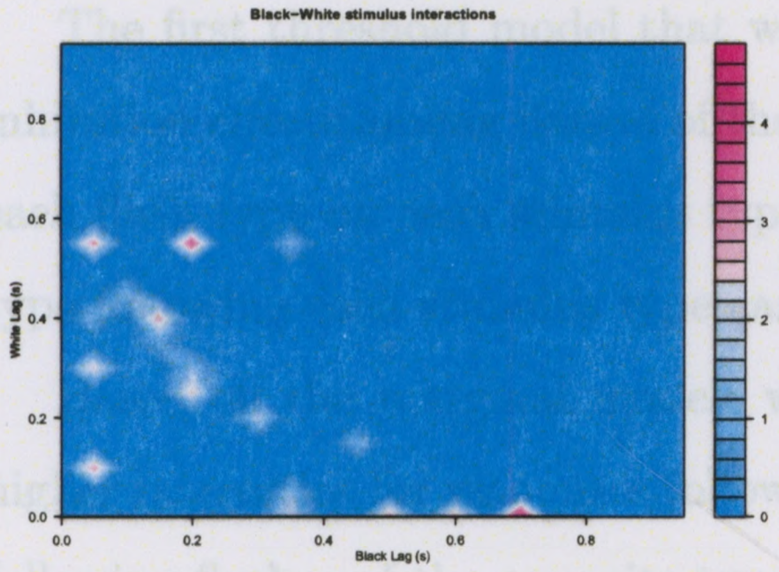


(a) White-White flash interactions

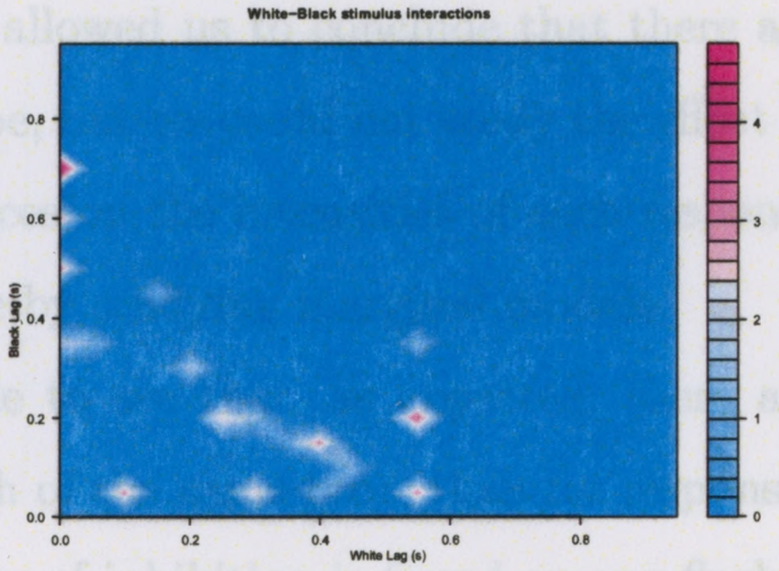


(b) Black-Black flash interactions

Figure 6.22: The second order intensities for the white responses of pooled choice RT data with rate $p_S = .5$.



(a) Black-White flash interactions

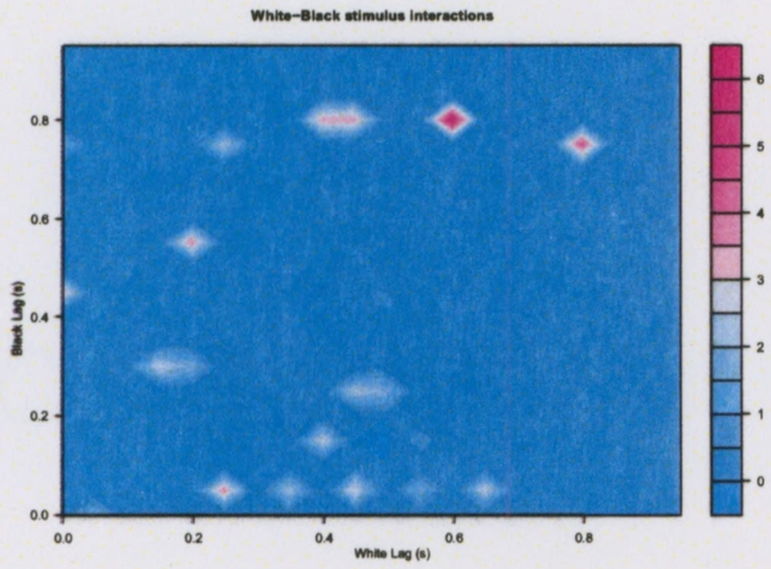


(b) White-Black flash interactions

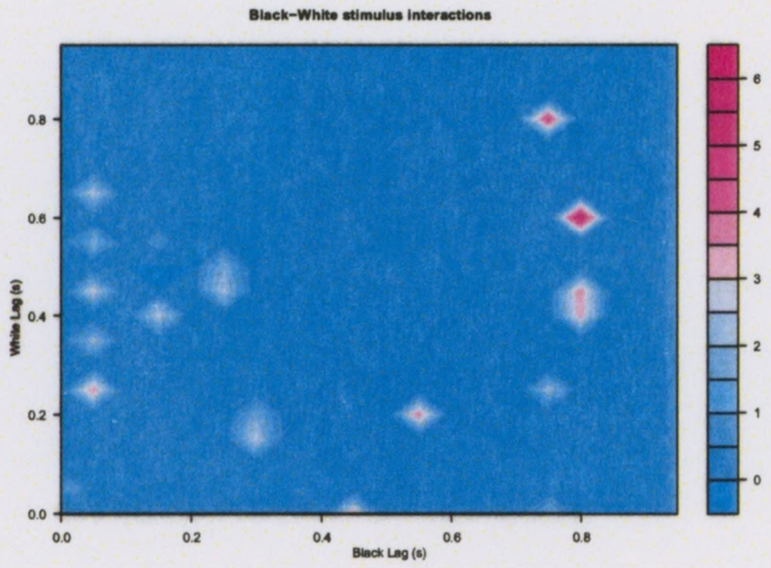
Figure 6.23: Second order intensities for the black responses of pooled choice RT data with rate $p_S = .5$.

6.3 Discussion

In this chapter we developed various threshold models for choice RT data and applied them to both real and simulated data. Fitting such models allowed us to obtain estimates of the threshold for each response type, as well as estimate certain impulse responses. Threshold estimates for the real data sets with different flash rates suggest that as the flash rate increases, a lower rod current is required for a response to occur. Each of the threshold models studied gives us evidence that there is no bias in response type. That is, when the two stimuli types are equiprobable, the two response types occur with similar frequencies.



(a) White-Black flash interactions

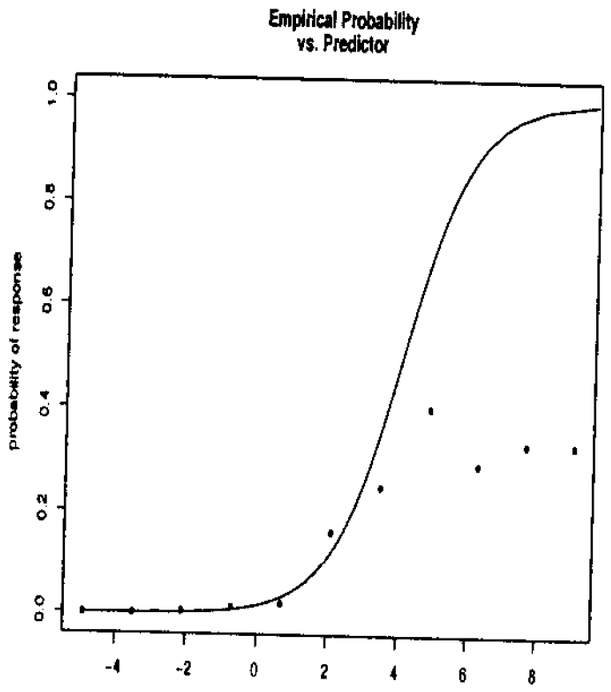


(b) Black-White flash interactions

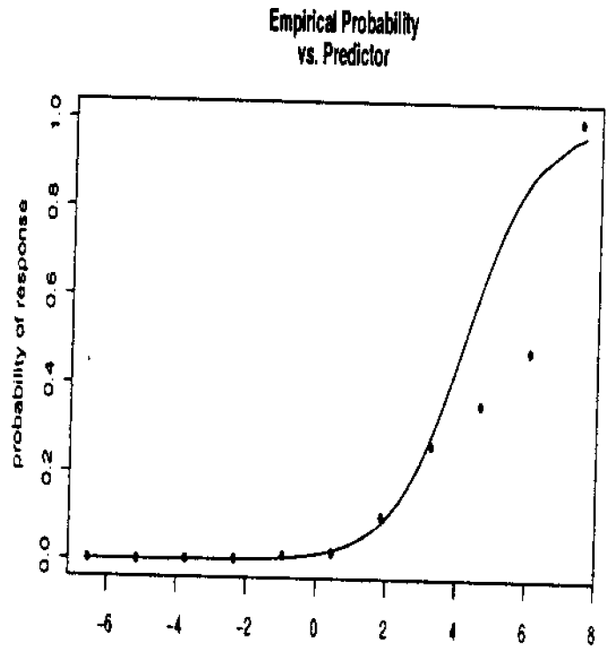
Figure 6.24: The second order intensities for the white responses of pooled choice RT data with rate $p_S = .5$.

The first threshold model that we studied allowed us to conclude that there are inhibition effects among flashes of the same type, but we could not study the effect of each flash type on each stimulus type. Inferences on the intensities of each response type following each stimulus type can be made by studying marginal models.

Based on the marginal models we are able to see that, as expected, there are higher intensities for responses following a flash of the same type, than for responses following flashes of the opposite type. Evidence of inhibition is found among flashes of the same type, and for lower lags there are a few cases of facilitation. In order to study interactions between flashes at different lags before a response type, and their effect on the response, a marginal model with all four interaction combinations of stimulus types at different lags is developed. Based on contour plots of these bivariate coefficients we are able to make inferences regarding any interactions that occur between flash types that are different lags before a response.



(a) Black Response Marginal Model



(b) White Response Marginal Model

Figure 6.25: Diagnostic plots for pooled choice RT data with rate $p_S = .5$.

Chapter 7

Conclusions and Future Research

In this work we introduce two types of models for reaction time data, and focus on three types of RT experiments. Each model type is used for a different purpose.

The parametric models that we develop are used in conjunction with nonparametric estimation to make inferences on the eye-brain-hand system. Similarities between nonparametric and parametric intensity estimates can indicate that common features in the nonparametric estimates are predicted by the parametric model. Furthermore, features unique to the nonparametric estimates suggest characteristics that the model cannot predict. For some simulated and real data sets, there is a peak in nonparametric estimates of response-response intensity functions involving the same response type, while our model always predicts a valley.

As an alternative parameter estimation method for runs of simple RT data with different flash rates, we propose a nonlinear regression method. Such a method is useful when it is assumed that the data sets have the same parameters, and this method has the advantage of a solid theoretical basis underlying hypothesis tests for the parameters. Based on the estimates obtained using the original estimation procedure, it appears reasonable to pool our simple RT data. Application of the nonlinear regression estimation to the simple RT data leads us to conclude that there is nonlinear inhibition present in the eye-brain-hand system, and that flashes are rarely deleted completely at random. Furthermore, our results indicate that there is negligible internal noise within the visual system. These results depend heavily on

the validity of our model.

Our fitted go-no go parametric models suggest that when flashes are presented at a fast rate (e.g. more than .5/s for each flash type), flashes that are close together in time are more likely to be perceived as one bright flash (i.e. temporal summation occurs). As in the simple RT experiment, our results imply that there are few occurrences of completely random thinning. Estimates of the error response probability (responding to a white flash), tend to be around 0.03.

In our fitted choice RT models, temporal summation is more likely to occur at both slow and fast flash rates, but the lag between flashes can be larger for faster flash rates. Our results suggest that at lower flash rates, completely random thinning appears to occur, but is relatively rare for faster flash rates. Estimates of the error probabilities (responding incorrectly to a flash) are at least 0.03, and seem to increase with the flash rate.

Estimates of the mean reaction time increase with the complexity of the RT experiment. For simple RT experiments, the mean is near 0.3, for go-no go RT a button is pressed around 0.4 s after a flash, while the mean delay is 0.5 for choice RT. These results are consistent with previous observations (Luce (1986)).

The variables in threshold models have direct biological interpretations. Fitting a threshold model gives us a procedure for estimating the minimum rod current at which a flash type can be detected. For all three types of RT experiments, we find that threshold estimates decrease as the flash rate increases. Threshold estimates for the choice RT experiments suggest that there is very little bias in response choice. Coefficients in the model can be used to make inferences on the presence of inhibition or facilitation among the flashes. The location of the peak for the impulse response coincides with the mean RT, and our results agree with the estimates obtained for our parametric models.

In the fitted go-no go threshold models, interaction coefficients tend to be negative, suggesting that pairs of flashes are associated with a decrease in the response rate, so that inhibition is present among the flashes.

There is evidence of inhibition effects between flashes of the same type, in the

fitted choice RT threshold models. Fitted marginal threshold models also suggest that when there is a small delay between a flash and response (e.g. lag smaller than 0.3), pairs of flashes actually increase the response rate, indicating the presence of facilitation.

We also fit a random threshold model to runs of simple RT data with the same flash rate, as an alternative to pooling the runs. Our results indicate that the data sets are similar enough to use simple pooling.

7.1 Future Research

Some extensions of the present work might be considered for future research. I think that the most natural could be a relaxation of a simplifying assumption in our parametric models for go-no go and choice RT data. Since the assumption of identical delay distributions for responses from black and from white flashes may not always be the case, dropping this assumption may improve the proposed model.

When the delay distribution is symmetric, the location of the peak of stimulus-response nonparametric intensity estimates coincides with the corresponding mean delay. That is, in the case of a symmetric delay distribution, the location of the peak of $\hat{p}_{S_W R_W,h}$ can be used to estimate the mean reaction time from a white flash. Therefore, in that case, S-R nonparametric intensity estimates can be used to give support to the appropriateness of our assumption of identical delay distributions. However, based on the results of our simulations, if the peak locations of the S-R nonparametric intensity estimates are not the same, it does not imply that our assumption fails to hold. In the case of simulated data we know that the delay distributions are identical, but the locations of the peaks in S-R nonparametric intensity estimates do not always support this. If the delay distribution is not symmetric, then the peak location approximates the mode of the delay distribution, and the S-R intensities cannot be used to make inferences on the mean RT.

Under a generalization that allows different delay distributions for responses from black and from white flashes the second-order intensity function expressions derived

in Sections 4.4 and 5.4 are no longer valid. However, it is simple to modify the intensity expressions in order to accommodate different delay distributions. Similar arguments can be used in the derivations, and it is only necessary to replace $f(v)$ by the respective density $f_B(v)$ or $f_W(v)$ depending on whether the response comes from a black or white flash, respectively.

When different delay distributions are permitted parameter estimation becomes more difficult. Our method for estimating the mean and standard deviation of the single delay distribution does not extend to the case of two delay distributions. As crude estimates of the means, the peak locations for appropriate stimulus-response intensities may be used. However, corresponding standard deviation estimates cannot be found by iterating through the data because it is impossible to match up the black flashes with possible white and black responses, and pair white flashes with possible white and black responses while simultaneously keeping the appropriate delay between flashes and responses; the delay between flashes and responses changes with flash type. Thus, alternative estimation methods should be sought.

References

Asimit, J. and W. J. Braun (2005). Application of third order intensity function estimation to reaction time experiment data. *Canadian Journal of Statistics 33*, 243–258.

Braun, W. J., V. Rousson, W. A. Simpson, and J. Prokop (2003). Parametric modelling of reaction time experiment data. *Biometrics 59*, 661–669.

Brillinger, D. (1975a). The identification of point process systems. *Annals of Probability 3*, 909–924.

Brillinger, D. (1975b). Statistical inference for stationary point processes. *Stochastic Processes and Related Topics 1*, 55–99.

Brillinger, D. (1976). Identification of synaptic interactions. *Biological Cybernetics 22*, 213–228.

Brillinger, D. (1988a). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics 59*, 190–200.

Brillinger, D. (1988b). The maximum likelihood approach to the identification of neuronal firing systems. *Annals of Biomedical Engineering 16*, 3–16.

Cornsweet, T. (1970). *Visual Perception.* New York : Academic.

Daley, D. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes* (second ed.), Volume 1: Elementary Theory and Methods of *Probability and its applications.* Springer.

Deutsch, S. and A. Deutsch (1993). *Understanding the Nervous System.* New York : IEEE Press.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology 39*, 197–215.

Genova, B., S. Mateeff, C. Bonnet, and J. Hohnsbein (2000). Mechanisms of simple and choice reaction to changes in direction of visual motion. *Vision Research 40*, 3049–3058.

Heinrich, L. (1988). Asymptotic gaussianity of some estimators for reduced factorial moment measures and product densities for stationary poisson cluster processes. *Statistics 19*(1), 87–106.

Jacod, J. and P. Protter (1991). *Probability Essentials.* Springer, Berlin.

Kreegipuu, K., C. Murd, and J. Allik (2006). Detection of colour changes in a moving object. *Vision Research 46*, 1848–1855.

Lahey Computer Systems, Inc. (1999). *Lahey/Fujitsu Fortran 95.* FUJITSU, LTD.

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

Montgomery, D. C., E. A. Peck, and G. G. Vining (2001). *Introduction to Linear Regression Analysis* (Third ed.). Wiley Series in Probability and Statistics. John Wiley and Sons.

Prokop, J. (2004). Application of third order intensity function to the analysis of reaction time experiment data. *University of Western Ontario Technical Report TR04-01*, 1–42.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Reiss, R. D. (1993). *A Course on Point Processes.* Springer-Verlag, New York.

Ripley, B. (2006). *Main Package of Venables and Ripley's MASS: an R Package.* R Foundation for Statistical Computing. S original by Venables & Ripley. R port by Brian Ripley ¡ripley@stats.ox.ac.uk¿, following earlier work by Kurt Hornik and Albrecht Gebhardt.

Simpson, W. A., W. J. Braun, C. Bargen, and A. J. Newman (2000). Identification of the eye-brain-hand system with point processes: A new approach to simple reaction time. *Journal of Experimental Psychology 26*, 1675–1690.

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Logitudinal, and Structural Equation Models.* Boca Raton : Chapman & Hall/CRC.

Stephen Wolfram (2006). *Mathematica for Students 5.2.* Wolfram Research.

Stoyan, D., W. S. Kendall, and J. Mecke (1987). *Stochastic Geometry and its Applications.* Wiley series in probability and mathematical statistics. Toronto : Wiley.

Ulrich, R. and J. Miller (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology 37*, 513–525.

Wand, M. P. and M. C. Jones (1995). *Kernel Smoothing.* Chapman and Hall.

Welford, A. T. (1980). *Reaction Times.* Academic Press.