

Modeling sequential context effects in judgment analysis: A time series approach

Jason W. Beckstead*
College of Nursing
University of South Florida

Abstract

In this article a broad perspective incorporating elements of time series theory is presented for conceptualizing the data obtained in multi-trial judgment experiments. Recent evidence suggests that sequential context effects, assimilation and contrast, commonly found in psychophysical judgment tasks, may be present in judgments of abstract magnitudes. A time series approach for analyzing single-subject data is developed and applied to expert prognostic judgments of risk for heart disease with an emphasis on detecting possible sequential context effects. The results demonstrate that sequential context effects do exist in such expert prognostic judgments. Contrast and assimilation were produced by cue series; the latter occurring more frequently. Experts also showed assimilation of prior responses that was independent of the cue series input. Time series analysis also revealed that abrupt or large trial-by-trial changes in the value of cues that receive the most attention in prognostic judgment tasks can disrupt the accuracy of these judgments. These findings suggest that a time series approach is a useful alternative to ordinary least squares regression, providing additional insights into the cognitive processes operating during multi-cue judgment experiments.

Keywords: expert judgment, time series, contrast and assimilation, single-subject analysis.

Psychological data are segments of life histories: as such they are ordered sequences of observations and by definition time series. — Robert A. M. Gregson (1983).

1 Introduction

Many judgment experiments may be viewed as involving two time series: a series of stimuli presented by the experimenter, and a series of responses provided by the subject. Over the years various theories of human judgment have been proposed. One thing all these theories have in common is that they have been, and are being, developed using single-subject repeated measures experiments. Whether judgment data are modeled using multiple regression, as is typical in the judgment analysis paradigm often associated with social judgment theory (Hammond, et al., 1975), by single-subject ANOVA which forms the foundation of information integration theory (Anderson, 1981; 1982), by conjoint analysis (Luce & Tukey, 1964; see also Krantz & Tversky, 1971), or by fast and frugal heuristics, such as Take the Best (Gigerenzer et al., 1991), or the Matching Heuristic (Dhami & Ayton, 1998; 2001), the data are obtained by presenting the subject with a series of stimuli to be judged

and recording a series of responses. Such idiographic designs (and analysis of their ensuing data) are the focus of this article.

Contrast and assimilation are psychological processes involving the sequential context in which judgments are made. In a variety of psychophysical judgment paradigms employing large numbers of trials (e.g., absolute and relative magnitude scaling tasks, absolute and relative identification tasks) sequential context effects are frequently observed (for reviews see DeCarlo & Cross, 1990; Stewart, Brown, & Chater, 2005). *Assimilation* occurs when the response to a given stimulus intensity tends to be larger when the immediately preceding stimulus is of greater intensity than the current stimulus, and tends to be smaller when the preceding stimulus intensity is less than that of the current stimulus. *Contrast* occurs when the response to a given stimulus intensity tends to be smaller when the immediately preceding stimulus is of greater intensity than the current stimulus, and tends to be larger when the preceding stimulus intensity is less than that of the current stimulus. DeCarlo and Cross (1990) discuss various theoretical models of psychophysical judgment that have been proposed to explain sequential context effects in magnitude scaling experiments and show how these can be evaluated using time series regression. One class of models is referred to as relative judgment models in which the subject is portrayed as com-

* Address: Jason W. Beckstead, University of South Florida College of Nursing, 12901 Bruce B. Downs Boulevard, MDC22, Tampa, Florida 33612. Email. jbeckste@health.usf.edu.

paring the value of the current stimulus to the value of the stimulus on the preceding trial rather than to some fixed internal reference. Stewart et al. (2005) discuss relative judgment models in absolute identification tasks. Another class of models suggests that sequential context effects result from a response heuristic, or the tendency of the subject, in the face of uncertainty, to guess in the direction of his or her previous response (see DeCarlo & Cross, 1990 for discussion). The point for consideration is that sequential context effects may arise because internal representation of the current stimulus is affected by the previous stimulus, and/or because of a tendency under uncertainty to provide a response based on the previous response. The question is whether similar sequential context effects operate in expert judgment tasks.

Expert prognostic judgments, such as a clinician's estimate of the likelihood that a patient will suffer a heart attack in the future based on current signs and symptoms, have been studied by judgment researchers employing multiple regression analysis (e.g., Beckstead & Stamp, 2007; Harries, 1995; Tape, Kripal & Wigton, 1992). To illustrate how sequential context effects might manifest themselves, let us examine what could happen with a single dichotomous cue. For example, consider the clinician faced with the prognostic task of estimating a patient's risk for coronary heart disease (CHD) and say that the cue in question is whether or not the patient has diabetes. The situation can be described by *assimilation* if the judgment of CHD risk for a patient with diabetes tends to be lower when the immediately preceding patient does not have diabetes, and, the judgment of CHD risk for a patient without diabetes tends to be higher when the preceding patient has diabetes. Alternatively, the situation can be described by *contrast* if the judgment of CHD risk for a patient with diabetes tends to be higher when the immediately preceding patient does not have diabetes, and, the judgment of CHD risk for a patient without diabetes tends to be lower when the preceding patient has diabetes.

In psychophysical judgment tasks, such as magnitude scaling, the observed response provided by the subject is interpreted to be an estimate of the sensory magnitude associated with a given, unidimensional, stimulus intensity. In multi-cue judgment tasks where there is not necessarily an objective stimulus intensity to be scaled, the response provided by the subject may be interpreted differently. In such judgment tasks, the observed response may be interpreted as an estimate of an integrated judgment along a more abstract subjective continuum, such as patient-risk for CHD.

Vlaev and Chater (2007) asked whether contrast and assimilation, as observed in psychophysical judgments, would operate when people make estimates of more abstract magnitudes. They examined estimates of cooperativeness made in a series of strategic choice decisions

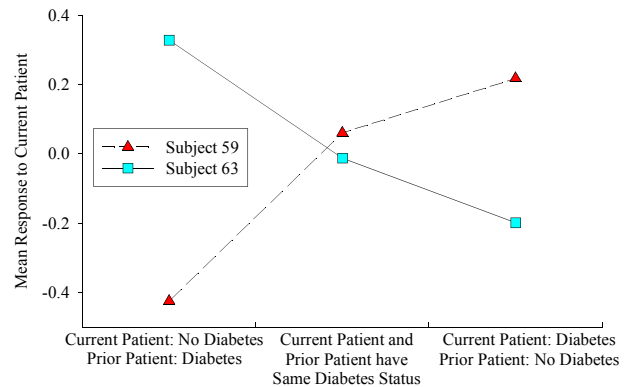


Figure 1: Sequential context effects produced by the diabetes cue in a multi-cue judgment task where clinicians judged patient's risk for coronary heart disease. Responses to the current patient (trial) are categorized according to consecutive values of the diabetes cue. Subject 63 shows assimilation, Subject 59 shows contrast. Plotted values are adjusted for category differences on the eight cues.

(i.e., prisoner's dilemma games). In their experiment, playing a random sequence of 96 cooperative and uncooperative games produced greater mean differences in cooperation rates (71% vs. 18%) when compared to conditions composed of 48 cooperative games followed by 48 uncooperative games (33% vs. 18%), and vice versa (18% vs. 50%). These differences were analyzed using ANOVA on aggregated responses. The authors interpreted the significant interaction as support for trial-by-trial (local) contrast effects. The current article investigates whether such sequential context effects operate in expert prognostic tasks but takes a different theoretical and analytical approach.

When sequential context effects associated with a cue in a multi-cue judgment task are observed, they are here interpreted to mean that a cue's *influence* (as represented by its β weight) is altered by the values that the cue takes over consecutive trials. Assimilation means that when the cue values on trials t and $t - 1$ are different, the cue's influence is *smaller* than when the cue values on these trials are the same. Contrast means that when the cue values on trials t and $t - 1$ are different, the cue's influence is *larger* than when the cue values on these trials are the same. Although this interpretation may sound odd to a psychophysicist, it is consistent with traditional methods of demonstrating sequential context effects. For example, one way of demonstrating these effects is to plot the mean response to the stimulus value on the current trial as a function of the differences between the stimulus value on the current and immediately preceding trial.

Figure 1 is an example of such a plot using data from two clinicians who participated in a judgment task

Table 1: Changes in influence weight, β , for the diabetes cue as a function of the cue's values on consecutive trials.

Subject	Effect	All trials	$\text{cue}_{(t)} = \text{cue}_{(t-1)}$	$\text{cue}_{(t)} \neq \text{cue}_{(t-1)}$
63	assimilation	.493	.575	.223
59	contrast	.404	.281	.788

Note: Number of trials total is 80, number of trials where consecutive cue values were equal is 54, number of trials where cue values were not equal is 25.

wherein they estimated 80 patients' risk for CHD based on eight cues (the task will be described in detail below). Here we focus on the diabetes cue for illustration. Each subject's responses were analyzed separately and converted to standardized scores to allow for direct comparison. Subject 63 (squares and solid line) shows assimilation; the mean rating of risk for the current patient is biased toward the diabetes status of the prior patient. Subject 59 (triangles and dashed line) shows contrast; the mean ratings are biased away from the diabetes status of the prior patient. Each subject's responses were also analyzed separately using multiple regression (diabetes was coded 1 if present, 0 if absent). Table 1 shows some of the results. First, when responses from all 80 trials were analyzed, both subjects shows roughly the same size partial β weights for the diabetes cue. When the data are categorized according to the values of the diabetes cue on consecutive trials and re-analyzed, we see how the β weight for each subject changes when assimilation or contrast takes place.

Although these regression analyses illustrate that the cue's influence changes when assimilation and contrast take place, this approach is flawed because of its piecemeal nature; data from two subsets of trials have been analyzed separately. A better approach would be to analyze the data from all the trials simultaneously.

Time series analysis can be used to test hypotheses that sequential context effects are operating in the series of responses obtained from a single subject. Specific models can be constructed to isolate sequential context effects produced by the cue (stimulus) series and those operating independently in the judgment (response) series. In the present article a time series approach is developed by extending ideas discussed in the context of psychophysical research to cover multi-cue judgment tasks. Before preceding to discuss the application of time series analysis, a broad perspective in which to position time series theory and methods in psychological research is needed. The next section is an attempt to provide such a perspective. Following this discourse, an illustrative application is presented with an emphasis on detecting possible sequential context effects operating in expert prognostic judgments of risk for heart disease.

2 Psychology from a time series perspective

2.1 An introduction to time series

A time series is a realization of a data-generating process, where observations are equally spaced across time. Familiar examples in econometrics include a stock's daily price, or quarterly sales figures (Yaffee, 2000). In terms more familiar to psychologists, Gregson (1983) defined a time series as a sequence of events ordered in time, which we may have good reason to believe is generated by some lawful underlying process that itself persists throughout the whole duration of the observations made. In psychology the series may be the responses a subject gives on successive trials of an experiment or the amount of some behavior a client undergoing psychotherapy exhibits daily over several weeks. The data collected in the psychological laboratory, or in field studies, are considered sampled segments from ongoing processes that are amenable to representation by univariate stochastic difference equations. Most measurements taken in psychology may be regarded as discrete realizations of continuous processes. The trial in a judgment experiment is conceptually taken as the unit throughout this article and the series of cue values and judgments are considered a discretely sampled data system.

Time series analysis is a set of regression-based methods for analyzing data ordered sequentially in time. The goal of the analysis is to identify patterns in the sequence of values, that is, to identify how the values are correlated with themselves but offset in time, in order to gain some insight into the underlying process(es) that generated the data. A series is decomposed into numerous potential components. One of these is a random process, referred to in the parlance of time series theory as a series of "shocks." Overlaid on these shocks are various possible patterns. Most obvious of these are trends over time (including linear and quadratically increasing and decreasing means). A second pattern is the lingering effects of earlier values in the series (i.e., an autoregressive or AR process), and a third is the lingering effects of earlier shocks (i.e., a moving-average or MA process).

These patterns are not mutually exclusive and all three may be found in a given time series. Readable introductions in econometrics are available (see Ostrom, 1990; Yaffee, 2000) and more rigorous mathematical treatment may be found in Hamilton (1994). A thought-provoking monograph, surveying the elementary theory of time series and indicating where and how its use can increase insight into psychological processes that extend through time, has been written by Gregson (1983). His treatise focuses on data obtained in the psychological laboratory and will be relied upon heavily throughout this article.

2.1.1 A psychology of organism-environment interactions in time

Brunswik (1952) advocated for a psychology of systems, and suggested that the proper subject matter for study by psychologists is the organism as it interacts with, and adapts to, objects in its changing environment. Adaptation is a sequential process, and as such, those seeking to understand it can benefit by applying time series theory and analysis. In tracing the history and thematic relations of psychology to other sciences, Brunswik (1956) pressed for aligning the methods and explicit theorizing of psychology with those of other disciplines he described as “already recognized as statistical on character.” Arguably, Brunswik may have recognized the importance of sequential context effects in adaptation and foreseen the relevance of their study for psychology. Indeed, his schematic representation of history concludes with the (at the time of his writing, unrealized) contributions of time series theory and analysis, citing the work of Wiener (1949). When discussing probabilistic prediction he mentioned autocorrelation as being useful. Major advances in the theory and mathematics of time series, now taken for granted, occurred after Brunswik’s death, notably the work of Kalman (1960) and Box and Jenkins (1970).

Researchers (Hammond, Hursch, & Todd, 1964; Tucker; 1964), working with Brunswik’s lens model, developed the lens model equation which quantifies and relates the cue-criterion relationships in the environment, the cue-judgment relationships, and the correspondence between judgments and criterion. As useful as the lens model equation has proved to be as a framework for conceptualizing the expert judgment process (see Stewart, 2001), in its current form it does not accommodate autocorrelation that may be present in the judgments and environment. Although beyond the scope of this article, it may be possible to modify the lens model equation to accommodate sequential effects by incorporating concepts from time series theory.

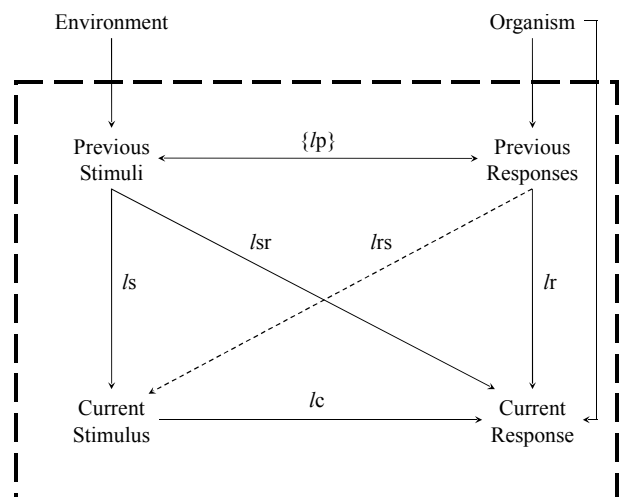


Figure 2: Dynamic structure of the organism-environment system.

2.1.2 Gregson’s dynamic structure of the organism-environment system

Gregson (1983) offered a framework for considering, and identifying, the dynamic structure of the organism-environment system using time series. He realized that the responses of an organism to its environment are not static and that adaptation may exhibit natural periodicity. He also recognized that the very act of doing an experiment in which responses are elicited to a series of stimuli can induce sequential dependency in responses. Like Brunswik, he recognized that the organism and the environment form a dynamic system. Gregson’s conceptual framework may be illustrated graphically (see Figure 2). The figure highlights the point that the study of organism-environment relationships is limited to measured stimulus-response relationships. Gregson treats the portion inside the double dashed line as a closed subsystem and regards it as the total scope of time series analysis. This closed subsystem contains all the quantitative data available to the researcher who wants to investigate organism-environment relationships.

Within this closed subsystem the various structures are considered in terms of their functional linkages, each of which may be the focus of one or more time series models. Using Gregson’s notation, we refer to the complete set of linkages $\{l\}$, and the component linkages are then:

l_c = current stimulus-response linkage.

l_s = linkage within the stimulus series.

l_r = linkage within the response series.

l_{sr} = linkage from previous stimuli to the current response bypassing the current stimulus.

l_{rs} = linkage from previous responses to current stimulus (this will be absent unless the stimuli are contingent

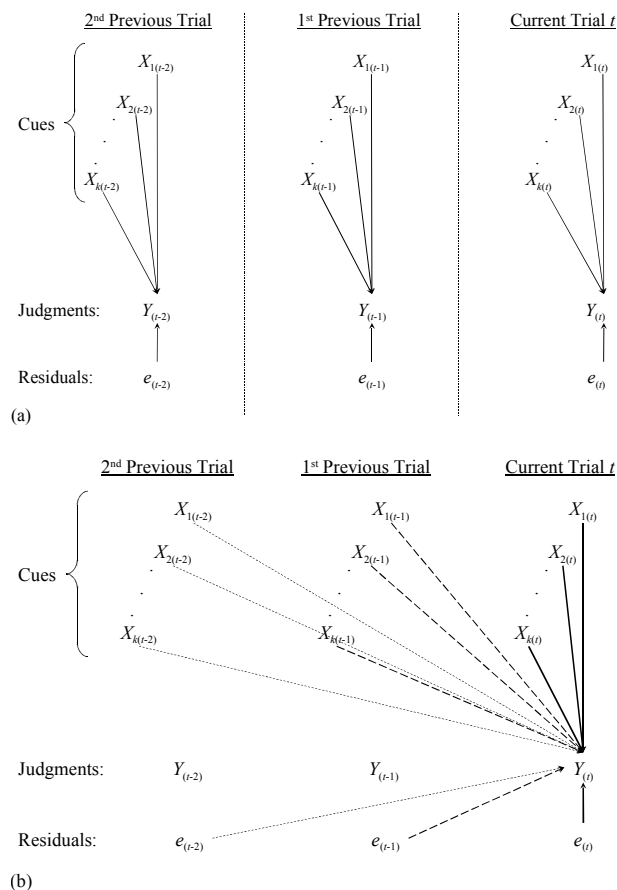


Figure 3: Schematic representations of multi-cue multi-trial judgment task. (a) Traditional view of judgment task (outside time). (b) Judgment task viewed from time series perspective. Note that the influence lines within previous trials are not shown in b for visual clarity.

upon previous responses; such linkage can exist if feedback has been introduced by the environment, which includes the actions of the experimenter).

$\{lp\}$ = past set of linkages, stimuli to responses and responses to stimuli.

Each of these linkages may be extant, or absent, in a time series model of current response (e.g., judgment) generation. Given the set of linkages $\{l\}$, the *general* problem of identification is to decide, using input-output records and notions of causal relations provided by psychological theory, which links are extant and which are absent. The *specific* problem of identification is one of deciding on the details of the algebraic structure and parameter values that most accurately represent what the links do, given that it is known which are extant.

Various experimental designs may be represented using $\{l\}$. In most psychological experiments the focus is on lc while ls is absent by design; the stimuli are presented in random order with the intent of reducing their

autocorrelation to zero. (Throughout the remainder of this discussion ls will be assumed absent by design.) In most psychological research the lr and lrs linkages are assumed (usually implicitly) to be absent. In many operant studies (and some judgment studies assessing the impact of feedback on accuracy) the investigator may be interested in lrs . In Figure 2 lrs is shown as a dotted line because, while the organism may derive feedback from the environment, the influence of such feedback cannot be investigated unless the value of the feedback stimulus is recorded from trial to trial. The $\{lp\}$ represents what the organism has learned through past interactions with its environment. For those readers familiar with Brunswik's writings, $\{lp\}$ may be analogous to what he called the texture of the environment.

The various linkages in Figure 2 correspond to basic time series models that may be applied to psychological data (i.e., to stimulus-response relationships) in general. If all linkages with the exception of lc are assumed absent, the model is considered to be outside time and involves no time series analysis. If responses are hypothesized to be generated by an autonomous process (i.e., a process that operates independently of the stimulus series), then only lr is assumed extant and the process is identified by an autoregressive (AR) structure. When the current response is hypothesized to be a function of current and previous stimuli, (lc and lrs are assumed extant; lr is assumed absent) the response-generating process is identified by a moving-average (MA) structure. More generally, when the current response is hypothesized to be a function of current and previous stimuli, as well as previous responses, all three types of linkage (lc , lrs , and lr) are assumed extant and the process is identified by an autoregressive-moving-average (ARMA) structure. "The general identification problem may be productively approached by assuming an ARMA structure and estimating the parameters within it or by seeking directly for a MA or AR solution; as the latter two are restricted forms of ARMA, this can eventually give the same result" (Gregson, 1983, p.27). The time series approach outlined by Gregson thus incorporates the principles from relative judgment models and response heuristic models; DeCarlo and Cross (1990) develop this idea in detail although they do not refer to Gregson's work.

2.1.3 Incorporating time series into judgment analysis

Consider a judgment experiment in which the subject is presented with a series of m profiles, each composed of k cues, and makes a series of m judgments. Each profile-judgment is here referred to as a trial, ranging from 1 to m . On the current trial t , judgment $Y_{(t)}$ is a function of the current cue values ($X_{1(t)} \dots X_{k(t)}$) and $e_{(t)}$,

representing residual or unmodeled sources of influence (see Figure 3a). Following Gregson, the solid lines are here referred to as *influence lines* representing the impact of the cues and that of the amalgamation of unmodeled sources. In regression-based judgment models (e.g., Hammond, Stewart, Brehmer & Steinmann, 1975) the strengths of the cue influences are often estimated using ordinary least squares multiple regression (OLSMR) coefficients and are assumed constant across trials; the influence of $e_{(t)}$ may be expressed as $1 - R^2$. In other models of judgment proposed by investigators working within the Brunswikian tradition (e.g., Dhimi & Ayton, 1998, 2001; Gigerenzer, Hoffrage & Kleinbölting, 1991), the strength of a cue's influence can vary from trial to trial. An assumption shared by all these models is that the cue values from previous trials do not influence judgments on the current trial. This assumption is represented by the vertical lines demarcating the trials in Figure 3a. In other words, these judgment models are outside time, limited to *lc* linkage; the *lr* and *lsr* are assumed absent. Subsequent discussion will be limited to regression-based models because they assume (at least initially) a constant cue influence throughout the series of trials and because they include a residual term ($e_{(t)}$) that is conveniently defined mathematically. These two qualities are important in the proposed time-series-analytic approach developed below.

When considered inside time, the same judgment experiment may be depicted as in Figure 3b. The *lc*, *lsr* and *lr* linkages are assumed extant; *ls* is absent by design. The lightening of the influence lines, from the current trial through the second previous trial, represents the weakening impact of prior cues and judgments occurring more distant in the series. Note that in regression-based models of judgment $Y_{(t)} = Y'_{(t)} + e_{(t)}$, where $Y'_{(t)}$ is the portion of the judgment that can be predicted from the cue values ($X_{1(t)} \dots X_{k(t)}$) and their regression coefficients, and $e_{(t)} = Y_{(t)} - Y'_{(t)}$ is the portion of the judgment that cannot be so predicted. As such, the series $e_{(t)}$, $e_{(t-1)}$, $e_{(t-2)}$, ... corresponds to *lr* and represents the influence of prior judgments with the effects of the cues partialled out via *lc*. This source of influence captures the response heuristic described above. To represent all these linkages and their relationships mathematically a type of time series model known as a linear transfer function model may be used.

A linear transfer function (LTF) model depicts the relationship between an output series and one or more input series. This class of time series models characterizes the autocorrelation function of the output series and the autocorrelation function of each input series (each of which is zero by design in most judgment experiments), as well as the cross-correlation functions between each input series and the output series. A cross-correlation function describes how lagged values of an input series are correlated

with the output series. For example, over the series of trials the correlation between the judgments and a cue's values, where the cue series is lagged by one trial, defines a first-order cross-correlation; when the cue series is lagged by two trials we have a second-order cross-correlation, and so on. For completion, the correlation between cues and judgments concurrent on the same trial is referred to as a zero-order cross-correlation. Linear transfer function models consist of two parts; the first part describes the relationships among the input and output series and the second part depicts the autoregressive structure of the residuals after cross-correlations have been fitted. Gregson's linkages (*lc*, *lsr*, and *lr*) may be elegantly represented in this class of time series models.

I propose that a linear transfer function autoregressive-moving-average model, incorporating general principles from psychophysical models is the best way to depict sequential context effects that may be operating in multi-cue judgment experiments involving several trials. Judgment is modeled as a function of the current cue values and the immediately preceding values of each cue. This relationship corresponds to a relative judgment model in psychophysics and is identified as a MA1 structure, where 1 refers to a first-order cross-correlation. The MA1 structure of the model uses the values of each cue on the current and immediately preceding trial throughout the series to provide parameter estimates of the extent to which the influence of the cue is modified by changes in its consecutive values.

The linear transfer function model also includes an AR1 error term to represent the portion of the judgment process that cannot be predicted from the cue values and their MA1 parameters. Inclusion of this AR1 structure accommodates the gist of the response heuristic model which suggests that people exhibit sequential context effects originating in their responses, independently of those operating in their perceptions of stimuli; or in Gregson's framework the *lr* linkage is assumed extant. The model is specified as:

$$Y_{(t)} = \mu + \sum_{i=1}^k (\beta_{0i} X_{i(t)} - \beta_{1i} X_{i(t-1)}) + \left(\frac{e_{(t)}}{1 - \varphi e_{(t-1)}} \right) \tag{1}$$

where $Y_{(t)}$ is the value of the judgment on the current trial, μ is the mean of the judgment series, $X_{i(t)}$ is the value of the *i*th cue on trial *t*, β_{0i} is a weighting coefficient for the *i*th cue on trial *t*, β_{1i} is a weighting coefficient for the *i*th cue on trial *t* - 1, $e_{(t)}$ is the residual on the current trial, $e_{(t-1)}$ is the residual on trial *t* - 1, and φ is an autoregressive weighting coefficient (limited to range from -1 to 1) for $e_{(t-1)}$.

In Equation 1 the influence the i th cue is conveyed using two parameters in order to represent possible sequential context effects. If the cue is used when forming a judgment then the sum of the absolute values of its two parameter estimates (β_{0i} and β_{1i}) will be greater than zero. If the cue produces contrast effects during the judgment process these two parameter estimates will have the same valence (note that β_{1i} has a negative sign in Equation 1). If the cue produces assimilation, the two parameters will have opposite signs. In DeCarlo and Cross's (1990) time series model, assimilation versus contrast is conveyed solely by the sign of β_{1i} because β_{0i} is always positive owing to the fact that sensory magnitude is positively correlated with stimulus intensity. In many multi-cue judgment tasks some cues will have negative correlations with values on the judgment dimension and so the signs of both β_{0i} and β_{1i} are necessary to distinguish contrast from assimilation effects. The magnitude of β_{1i} (positive or negative) estimates the extent to which the cue's influence changes due to differences in the cue's values on consecutive trials. This method of estimating the change in a cue's influence is more reliable than the piecemeal approach used in Table 1 because the estimate is based on data from all trials rather than subsets of trials.

Equation 1 also accommodates sequential context effects that may be operating in the response series independently of the cues. If the judge has a tendency, in the face of uncertainty, to guess in the direction of his or her previous response, this form of assimilation will result in a positive value of φ . If the judge tends to guess in the opposite direction (i.e., contrast) φ will be negative. Thus, the approach assumes lc , lsr , and lr linkages are extant (ls is absent by design) and the model provides the means for quantifying contrast and assimilation operating within the cue and judgment series.

As a proof of concept, Equation 1 was fitted to data from a sample of nurse practitioners who performed a prognostic judgment task, estimating patient risk for CHD. The goals of this application are (1) to establish the utility of Gregson's framework for studying the dynamic structure of the organism-environment system for examining prognostic judgments, (2) to show that the linear-transfer-function autoregressive-moving-average (LTF ARMA(1,1)) model can fit such data better than OLSMR, and (3) to demonstrate that sequential context effects exist in prognostic judgment tasks.

3 An illustrative application

3.1 Method

3.1.1 Subjects

Seventy-five nurse practitioners completed a prognostic judgment task in which they made estimates of risk for

CHD for 80 patient profiles. Four of the nurse practitioners were male. The average age was 48.2 (SD = 6.8). Most (81.3%) worked in primary care settings. On average, subjects had 8.9 years of practice experience (SD = 6.3).

3.1.2 Materials

Selection of Cues and Outcome Measure. The optimal set of risk factors for predicting CHD were identified by Anderson, Odell, Wilson, and Kannel (1991a) using a sample of 5,573 patients followed for over 12 years as part of the ongoing Framingham study of heart disease. The equation of Anderson et al. provided regression coefficients for eight patient characteristics: gender, age, smoking status, total cholesterol level, high-density lipid level (HDL), systolic blood pressure (SBP), and whether or not the patient has been diagnosed with diabetes or left ventricular hypertrophy (LVH). This "gold standard" for predicting CHD was published in the form of a clinical worksheet later that same year (Anderson, Wilson, Odell, & Kannel, 1991b). In the current study, judgments of patient risk for CHD were made using a 0% to 100% response scale.

Choice of Cue Values. A representative design was used to construct patient profiles for the judgment task. The risk-factor distributions reported by Anderson et al. (1991a; 1991b) were used to generate a population of cases with similar means, variances, and correlations among the eight risk-factor cues. Eighty cases were randomly sampled from this population and randomly ordered for presentation in the judgment task.

3.1.3 Procedure

The materials were presented to each subject in a booklet. Booklets contained a cover page describing the purpose of the study ("to understand how nurse practitioners form judgments of patient risk for CHD"), instructions for the judgment task, the series of patient profiles presented separately in tabular format, and a brief section requesting basic demographic information. Nurse practitioners were instructed to "Please read each profile carefully and make an assessment of the patient's risk for CHD within the next 10 years on a scale of 0% to 100%." Nurse practitioners were tested individually and in small groups in office and classroom settings. After obtaining informed consent, instructions describing the judgment task and accompanying materials were read aloud to subjects. The procedure took an average of 32 minutes (SD = 11.6) to complete. All subjects received the profiles in the same order.

3.1.4 Preliminary analyses

Multiple-cue judgment analyses typically use standardized coefficients as estimates of a subject's cue weighting strategies. Prior to fitting the model in Equation 1, all variables (cues and judgments) were standardized for each subject individually to have means of 0 and variances of 1. This was done as a matter of interpretational convenience providing a common metric upon which to compare parameters across individuals and also because SAS's PROC ARIMA does not provide standardized parameter estimates. Second, for comparative purposes, each individual's judgments were regressed onto the cues using ordinary least squares multiple regression (OLSMR) and residuals were examined for serial dependence using two tests described in the next paragraph. Third, for each individual OLSMR was used to confirm that the functional relationships of all cues to judgments were linear (i.e., there were no quadratic trends in cue-judgment relationships). Fourth, the autocorrelation function for each cue was assessed to confirm that *Is* linkages were absent by design. Fifth, Dickey-Fuller tests (Dickey & Fuller, 1979) were used to confirm that all judgment series were stationary prior to fitting the time series model. Stationarity refers to a series having a constant mean and variance.

There are various methods for assessing serial dependence. One of the more well known methods is the Durbin-Watson test (Durbin & Watson, 1950; 1951). Cooksey (1996) discusses how the Durbin-Watson (DW) test may be applied in judgment analysis. An advantage of the DW test is that it is commonly available as an optional test of residuals in OLSMR procedures of most statistical packages (e.g., SAS and SPSS). A disadvantage is that it does not assess serial dependence beyond first-order autocorrelation. A second disadvantage of the DW test is that the derivation of its standard errors (and hence, critical values) is not straightforward. Cohen et al. (2003) discuss the DW test in detail. An alternative, and in the present application more useful, test is the Ljung-Box statistic (Ljung & Box, 1978) that can be used to assess a series for departures from "white noise" by simultaneously examining autocorrelations over a range of predetermined orders. The LB test is a weighted sum of squared autocorrelations. One criterion for identifying a correct time series model is that serial dependence in the residuals is reduced to zero (i.e., a white noise process). The LB test was developed as a means to make such assessments. The LB statistic is distributed as χ^2 where a nonsignificant result indicates the series is free from serial dependence or does not differ significantly from a white noise process. Based on the LB test, 26 individuals exhibited serial dependence; the DW test identified only 15 of these. Thus, it appears that when applied to

OLSMR residuals from a single subject, typical in judgment analysis, the LB test may be more sensitive for detecting serial dependence.

3.2 Results

3.2.1 Goodness of fit: Comparing OLSMR and LTF ARMA(1,1) on the basis of their R^2 values

The LTF ARMA(1,1) model was successfully fitted to 68 of the 75 subjects. The other seven required higher-order autoregressive terms to identify the AR portion of their judgment models and render the residuals as white noise. These will be discussed later. Except where mentioned explicitly, the remainder of this section focuses on the analyses from 68 individuals. The model was fitted using SAS's PROC ARIMA; parameters were estimated using conditional least squares rather than maximum-likelihood because this method has been shown to perform better when the number of trials is less than 100 (Yaffee, 2000, pp. 192–204).

R^2 values ranged from .552 to .928 with an average of .800. F tests revealed that the R^2 value for each subject was significantly larger ($p < .05$) than his or her OLSMR R^2 value (these ranged from .490 to .908 with an average of .750). These tests took into account the differing number of parameters between the two models. A second method for comparing the fit of these two models focuses on serial dependence in their residuals. The results of these analyses are reported below when discussing tests for serial dependence.

3.2.2 On determining whether a cue is used when forming a judgment

Reliance on tests of significance when determining whether a cue is being used by an individual in a judgment task has been recently called into question (Beckstead, 2007). An alternative to significance tests for determining whether or not a cue is influential is to focus on effect sizes. In the current application if a change of one standard deviation in a cue's value produced at least a .333 standard deviation change on the judgment scale, the cue was considered to have been used by the subject. Although arbitrary, this definition is somewhat conservative when considered in the context of traditional notions of effect size (see Cohen, 1988). As each cue was represented by two parameters (β_{0i} and β_{1i}) the sum of the absolute values of its two parameter estimates had to be $\geq .333$. For purposes of illustration, in order to have been considered as exhibiting contrast or assimilation the absolute value of each parameter estimate had to contribute at least .111 to this sum. Using these operational definitions, the distribution of cue utilization was as follows: five individuals used only one cue, 19 used two, 26 used

Table 2: Number of Individuals Exhibiting Sequential Context Effects on Each Patient Characteristic (Cue) used in Prognostic Judgments of Coronary Heart Disease.

Cue	Used	Contrast	Assimilation	No SCEs	Total
Gender	no	0	0	64	64
	yes	1	0	3	4
Age	no	0	0	50	50
	yes	2	1	15	18
SBP	no	0	0	40	40
	yes	1	13	14	28
LVH	no	0	0	43	43
	yes	0	4	21	25
Cholesterol	no	0	0	48	48
	yes	2	3	15	20
HDL	no	0	0	61	61
	yes	1	2	4	7
Smoking	no	0	0	28	28
	yes	1	2	37	40
Diabetes	no	0	0	11	11
	yes	4	1	52	57

Note: To be counted as being used in judgment, a one standard deviation change in the cue's value had to produce a .333 standard deviation change on the judgment scale. SCE = sequential context effect, SBP = systolic blood pressure, LVH = left ventricular hypertrophy, HDL = high-density lipids.

three, 12 used four, and six used five. For each subject who used multiple cues, the cues used were rank ordered according to their influence (1 being assigned to the most influential cue). This was done in order to ascertain if sequential context effects occurred more frequently as cues carried more influence.

3.2.3 Sequential context effects produced by cue series

Each cue was used by at least four of the subjects (see Table 2). The most frequently used cue was whether or not the patient had diabetes, and the least frequently used was patient gender. Each of the cues produced assimilation, contrast, or both, although with varying frequency among subjects. The blood pressure (SBP) cue produced sequential context effects, notably assimilation, for the majority of subjects who used the cue. Why this should

be the case is not clear. Although purely speculative, it is possible that nurse practitioners were more familiar with the range of values of this cue in relation to heart disease (perhaps it represents a defining characteristic) and that this induced stronger memory traces or increased processing of the information provided by the cue which ultimately manifested as assimilation from one trial to the next. Inspection of cue influence rankings revealed that sequential context effects tended to be more common with higher ranks, that is, they occurred more often for cues that carried more weight in the judgments. (See Table 3.) In total there were 38 instances of sequential context effects produced by the cue series (26 instances of assimilation and 12 instances of contrast). These instances were distributed among 30 individuals. Eighteen subjects displayed evidence of assimilation effects only, eight exhibited evidence of contrast effects only, and four showed both assimilation and contrast produced by different cues.

3.2.4 Sequential context effects in judgments independent of cue series

The LTF ARMA(1,1) model incorporated a parameter for quantifying the autoregressive structure of the residuals, that is, the degree of serial dependence in the judgment series that was independent of the cue influences. The results from 19 subjects included significant φ parameters ($p < .05$). These parameter estimates ranged from .264 to .766 with a mean of .433 indicating that assimilation (not contrast) was operating autonomously in the responses. When both β and φ parameters were considered, 39 subjects exhibited evidence of sequential context effects in the judgment task. For nine of these individuals this was limited solely to assimilation effects in the response series. Twenty exhibited sequential context effects produced by only the cue series, and 10 displayed evidence of sequential context effects operating in both cue and response series.

3.2.5 Comparing OLSMR and LTF ARMA(1,1) on the basis of serial dependence in their residuals

Using the DW test, only 12 of the 39 subjects who showed any form of sequential context effects in the time series analysis screened positive for serial dependence in their OLSMR residuals. The test missed 24 of the 30 who exhibited sequential context effects produced by the cues series and missed seven of the 19 who showed assimilation to previous responses. The DW test produced no false positives. As noted above, the LB test appears more sensitive than the DW test for detecting serial dependence in OLSMR residuals. Sixteen of the 39 subjects who showed sequential context screened positive for se-

Table 3: Frequencies of sequential context effects produced by cue series according to influence rank.

Rank	Cue							
	Gender	Age	SBP	LVH	Cholesterol	HDL	Smoking	Diabetes
1	0/0	2/7	3/3	2/5	1/6	2/2	1/8	4/37
2	0/1	0/6	6/11	0/8	3/8	1/2	1/17	1/10
3	1/2	0/2	3/10	0/7	0/2	0/2	1/12	0/7
4	0/1	0/2	2/4	1/3	0/3	0/1	0/1	0/3
5	0/0	1/1	0/0	1/2	1/1	0/0	0/2	0/0

Note: Cues were rank ordered according to the size of their parameter estimates (largest assigned rank of 1). Denominator is number of times cue appeared at each rank. Numerator is number of times cue exhibited sequential context effect. SBP = systolic blood pressure, LVH = left ventricular hypertrophy, HDL = high-density lipids.

rial dependence in OLSMR residuals using the LB test. The test missed 21 of the 30 exhibiting sequential context effects produced by the cues series and four of the 19 who showed assimilation to previous responses. The LB test produced no false positives.

Given the higher sensitivity of the LB test, it was used to make comparisons between the two analytic approaches. Applying the LB test to the residuals from the LTF ARMA(1,1) analyses for the 16 subjects with serial dependence in their OLSMR residuals revealed that LTF ARMA(1,1) residuals were rendered as white noise for all these subjects. Thus using serial dependence in residuals as the criterion, it appears that the LTF ARMA(1,1) model fit the judgment data better than the OLSMR model did.

3.2.6 Identifying higher-order autocorrelated structures in judgments

The remaining seven of 75 individuals exhibited persistent serial dependence in their data after fitting the LTF ARMA(1,1) model based on LB tests. Although atypical, such findings are not without precedent. Early empirical evidence (Holland & Lockhead, 1968) in the psychophysical realm suggested that autocorrelations up to eighth-order may be operating in some serial judgments. Later computer simulations by Gregson (1976) suggested, however, that first- and second-order processes are more psychologically plausible and that such higher-order findings were likely the result of model misspecification.

Higher-order AR structures in the absence of all intervening lower-order ones are known as periodic, seasonal, or cyclic. For example, in economic models of data recorded monthly, it is common to observe AR(12) structures that reflect the monthly cyclicity in sales or

spending patterns over several years (i.e., December data from year t are correlated with December data from year $t - 1$, January data from year t with January data from year $t - 1$, etc.). Although purely speculative, in self-paced judgment tasks, like the one examined here, some individuals may experience waxing and waning attention/concentration on the task from trial to trial, or they may engage in self-monitoring efforts producing alternating response set bias. These cognitive processes might possibly manifest as higher-order periodic AR structures. Of course, such higher-order structures might also be the result of unknown influences associated with conditions of the experimental context.

Identifying such higher-order autoregressive structures is largely an exploratory process. One exploratory approach uses SAS's PROC AUTOREG employing its backstep option to test the effects of including higher-order autoregressive parameters on reducing serial dependence in the residuals. This option removes nonsignificant autoregressive parameters (analogous to backward elimination in multiple regression) using Yule-Walker equations. (See Brocklebank & Dickey, 2003 for mathematical details.) What remains in the model is the most parsimonious autoregressive structure that accurately (within predefined limits) fits the data. This approach was used on the data from these seven individuals, testing for first- through eighth-order autoregressive parameters. This exploratory process identified idiosyncratic higher-order AR structures (Table 4). Three subjects (10, 12, and 42) showed evidence of pure periodicity; the others displayed more complex structures. Despite these atypical and atheoretical AR structures, the majority of these individuals showed evidence of sequential context effects produced by the cue series; three showed only cue assimilation effects, two showed only

Table 4: Idiosyncratic higher-order AR structures for seven atypical individuals.

Subject:	8	10	12	31	42	62	75
AR structure:	1,3,5	4	5	1,3	3	1,6	2,5,6
SCEs:	none	3c,1a	1c	1a	2a	1a	1c

Note: AR = autoregressive error term in linear transfer function autoregressive moving average model; SCE = sequential context effects produced by cue series, c = contrast, a = assimilation, and the values preceding these letters are the number of cues that produced each type of effect.

cue contrast effects, one showed both cue assimilation and contrast effects, and one showed no sequential context effects produced by the cue series.

3.2.7 A supplemental lens model analysis involving time series

In judgment studies employing the full lens model, the accuracy or achievement of the subject is defined relative to the environment in which the judgments take place and is quantified by the correlation of judgments with criterion values across the set of trials. The lens model does not represent accuracy on a trial-by-trial basis. If the lens model is considered as a dynamic system in time, where cue series represent input and the series of judgments represents output, then questions about the *stability* of the subject's accuracy may be addressed. For example one might ask, do abrupt or large changes in the cue series disrupt trial-by-trial accuracy of the judgment process? Applied in this context, time series analysis can be used to obtain insight into the how the judgment process adapts (on a trial-by-trial basis) to changes in the environment.

In the present application, the criterion for accuracy was CHD risk as calculated by the Framingham equation (see Anderson, et al., 1991a). To address the question of stability in accuracy, the eight cue values for each profile were entered into the Framingham equation to obtain the "correct" answer for each trial, $Y_{(t)}^*$. The error in judgment on each trial was then defined by taking the absolute value of the difference in Framingham risk estimate and the judgment of patient risk provided by the subject. Each cue series was *differenced* and then used to predict errors in judgment using a standard multiple regression analysis. Differencing a series refers to subtracting the value on trial $t - 1$ from value on trial t in order to form a new series in which the values represent the amount of change in the values of the original series from one trial to the next. Stating the question more formally in terms of time se-

ries: does $\nabla_{X_i} \rightarrow E_{(t)}$, where $\nabla_{X_i} = |X_{i(t)} - X_{i(t-1)}|$ and $E_{(t)} = |Y_{(t)}^* - Y_{(t)}|$? Furthermore, if such disruptive effects are observed, do they occur more frequently among cues that carried more influence in the judgment task?

In this separate analysis, 23 of 68 subjects showed evidence that changes in the cue series that they relied upon when making risk judgments disrupted their judgment accuracy. At least one cue had a significant ($p < .05$) effect on accuracy for each subject. The accuracy of three subjects was disrupted by two cues, yielding 26 instances of disruption. Twelve of these were produced by the highest ranked cue and five by the second highest. Thus it appears that abrupt or large trial-by-trial changes in the value of cues that receive the most attention in prognostic judgment tasks can disrupt the accuracy of the judgment process.

4 Discussion

In the application above, a linear transfer function autoregressive-moving-average model (LTF ARMA(1,1)) was successfully fitted to expert prognostic judgments of CHD risk made by 68 of 75 nurse practitioners. The general identification of the model was theoretically anchored within Gregson's (1983) framework for studying the dynamic structure of the organism-environment system. This framework proved useful for conceptualizing the correspondence between the various linkages (*lc*, *lsr*, and *lr*) assumed extant in the judgment process and the model's parameters. The specific identification of the model, that is the hypothesized MA1 structure for capturing sequential context effects produced by the cue series, and the hypothesized AR1 structure for accommodating assimilation in the response series were based on principles taken from magnitude scaling research in psychophysics (DeCarlo & Cross, 1990). The analysis of differenced cue series as predictors of trial-by-trial accuracy revealed that abrupt or large changes in the cue values can disrupt the stability of judgment accuracy. These applications demonstrated how time series analysis can be productively incorporated into the study of expert judgment.

Analysis of expert judgments is often conducted using ordinary least squares multiple regression (OLSMR). This study compared OLSMR to a time series approach for analyzing the data from a prognostic judgment task involving risk of coronary heart disease. Two criteria were used to evaluate the performance of these two analytic approaches. First, based on comparisons of coefficients of determination (R^2) the time series model fit significantly better for all subjects providing support for its ability to more accurately represent such judgments.

The second basis for comparing the goodness of fit for the two approaches centered on each models' ability to remove serial dependence from the residuals. For the subset of individuals identified by the LB test as showing serial dependence in their OLSMR residuals, the time series model effectively reduced each subject's residual series to white noise. Together these results support the LTF ARMA(1,1) time series model as an accurate and informative method for analyzing single-subject data from multi-trial judgment experiments.

To the best of my knowledge, this is the first article to examine sequential context effects operating in expert judgments, although Laming (1995) discusses similar ideas in the context of a retrospective analysis of cervical cancer screening by one expert. The analyses reported here revealed the presence of sequential context effects in judgments made by for 45 of 75 individual experts. Although these effects were not universal, the fact that they were shown by *any* individuals demonstrates that such effects can occur in prognostic judgments. (It is worth noting here that if a more liberal operational definition of cue usage, .250 rather than .333, were to have been used, the number of subjects exhibiting any sequential context effects would have totaled 62 not 45.) One possible process explanation for these sequential context effects involves the formation and influence of memory traces from one trial to the next, although the analyses reported here cannot be used to support or refute such an explanation. The findings do suggest that expert prognostic judgment tasks can provide another experimental paradigm in which to test the generality of such explanations.

When discussing time series modeling of stimulus-response relationships in the realm of psychophysical judgments, Gregson (1983) notes that is not unexpected to find different subjects showing evidence of AR1, AR2, MA1, MA2 and ARMA models operating in the same judgment experiment. In the current application of time series analysis to expert judgments, a confirmatory rather than exploratory approach was taken; a single model, theoretically based on relative judgment models and response heuristics, was fitted to all subjects and the degree and frequency of fit examined. The model was successfully fitted to 91% (68 of 75) of the sample of experts examined here. Some subjects showed evidence of only relative judgment processes operating in the cue series, some showed evidence of only a response heuristic operating, some showed evidence of both relative judgment and response heuristic influences, and some showed no evidence of sequential context effects at all. Future studies might examine personality variables and other individual difference factors that differentiate among people who exhibit various sequential context effects on expert judgment tasks and those who do not. It would also be interesting to explore key features of the judgment task,

such as the amount of environmental uncertainty, as a moderator of these effects.

It is possible that the observed sequential context effects may have been induced by the experimental context. Presenting "paper patients" on successive pages of a booklet may have drawn attention to the values of preceding cues. This in turn may have caused some individuals to form memory traces of preceding cue values which would otherwise not have been incorporated into subsequent judgments. If this interpretation proves to be correct the current application highlights the value of the proposed time series model for isolating such laboratory artifacts.

Another way to view the results is that the experimental context was *sufficient* to uncover sequential context effects operating in several expert judges. It is possible that presenting patient cue profiles in the experimental setting, stripped of the accompanying social aspects of the patient encounter, could have diminished the salience of the cue values and the extent of their cognitive processing. This in turn may have dampened sequential context effects produced by the cue series. If so, it is reasonable to expect sequential context effects to appear with greater intensity and frequency in the clinical setting where the salience of the information provided by the cues is increased, and the clinician's memory of preceding patient encounters is perhaps stronger.

It is not being suggested that experts use a LTF ARMA(1,1) model in their heads when making repeated judgments. What is being suggested is that, if one accepts the proposition that people's judgments can be modeled as though they are multiple regression equations, then OLSMR may be insufficient to capture the complexity of cognitive processes involved, because such models exist outside of time. Sequential context effects commonly observed in studies of psychophysical judgment and recently reported in strategic decision making do appear to exist in expert prognostic judgments. Contrast and assimilation were produced by cue series; the latter occurring more frequently in the task examined here. Experts also showed assimilation to prior responses that was independent of the cue series' influence. That is, when faced with uncertainty they tended to guess in the direction of their previous response.

In the current study all subjects received the series of 80 profiles in the same (random) order. This was done intentionally to isolate individual differences in cognitive processing for comparison. For instance, when two subjects showed different weights for a given cue, we know that this result was not due to differences in the order in which the profiles were experienced. Similarly, when one subject showed assimilation (or contrast) on a particular cue and the other subject did not, we can rule out differences in the profile order as producing this result. Had

each subject experienced a unique (and random) ordering of the profiles, this uniqueness would have been completely confounded with individual differences in the sequences of their responses.

Much has been written in psychology about how to construct and analyze repeated-measures experiments involving multiple subjects in order to minimize unwanted sequential effects or serial dependence (e.g., Keppel, 1991; Kirk, 1995; Myers, 1979; Maxwell & Delaney, 2004). Counter-balancing the order of treatments or stimuli across subgroups or randomizing order for each subject are often proposed as solutions when data are analyzed in aggregate. These solutions do not offer much solace to judgment analysts who conduct *single-subject repeated-measures analyses* however. When single-subject multiple regression analysis is conducted on judgment data, as is typically the case in social judgment theory paradigm (Brehmer & Joyce, 1988; Stewart, 2001) it seems, at least in principle, that a fundamental assumption of the regression model has been violated; the assumption that the data (more formally the residuals from the regression analysis) are independent does not appear tenable. This putative nonindependence stems from serial dependence or autocorrelation.

Serial dependence has two adverse consequences: (1) the standard errors for the regression coefficients are too small, leading to increased type I error when testing their significance (although estimates of the coefficients themselves tend not to be biased), and (2) the R^2 expressing the goodness of fit for the regression equation is biased. Some intuitive understanding for why the first consequence occurs may be gained by considering that the successive observations are dependent to some extent; thus, they provide less information than the same number of independent observations would. For example, if one had 100 actual observations, but because of serial dependence in the series there was only the equivalent of 50 independent observations worth of information, then 50 should be used in computing the standard errors but 100 gets used because the dependencies are ignored. This loss in efficiency of the OLSMR estimators for various degrees of autocorrelation is illustrated by Johnston (1984, pp. 310–313). The reason for the second consequence is a bit more complicated to explain because R^2 can be inflated or deflated depending on whether the serial dependence in the residuals is accompanied by serial dependence in the predictor variables, and whether these autocorrelations are of the same or different valences. In psychophysics investigators often generate many long series of random stimuli, assess the autocorrelation in each and retain only those series that are free from serial dependence for use in their experiments. DeCarlo & Cross (1990, p. 387) show that, under such conditions, correlated residuals add to the size of the error variance and hence produce a de-

flated estimate of R^2 . (It is not common practice among judgment analysts to assess the series of cue values used in judgment tasks for serial dependence; simply randomizing the order of a set of profiles does not guaranty an autocorrelation of zero). Wonnacott & Wonnacott (1979, pp. 206–208, 212–215) demonstrate that when there is a positive autocorrelation in both the residual and predictor series, the estimated regression line will fit the data very well, leaving small residuals and thereby inflating the estimate of R^2 .

These consequences may be of special interest to researchers working with Brunswik's lens model who place substantive interpretation on the value of R_s as a measure of the subject's cognitive control during the judgment task, and to judgment researchers, in general, who often rely on regression-weight significance tests to determine the number of cues used by the subject (see Beckstead, 2007). By incorporating time series theory and analysis into modeling human judgment the problematic statistical conditions which result from violating the independence assumption can be eliminated and additional insight into the nature of cognitive processes that influence sequential judgments can be gained. An alternative method, suggested by one reviewer, is to analyze data in aggregate using multi-level modeling techniques that incorporate an autocorrelated error structure. While this approach can handle the independence violation, it may obscure interesting individual differences such as when different subjects show assimilation or contrast to the same cue series.

Time series analysis, coupled with Gregson's framework for studying the dynamic structure of the organism-environment system, is a powerful alternative for analyzing human judgment, in keeping with the Brunswikian tradition. Time series theory holds potential for extending the lens model equation to accommodate dynamic aspects of the organism-environment system such as the disruption of trial-by-trial accuracy by abrupt changes in cue values.

Many theories of human judgment have been, and are being, developed using experiments that involve presenting the subject with a long series of stimuli to be judged and recording a series of responses. As such, there exists the possibility that cognitive processes involved in responding to sequential stimuli, but not necessarily limited to contrast and assimilation, could be operating within the individuals studied in these experiments. Whether or not these cognitive processes are activated under natural conditions as part of an organism's adaptation to its changing environment is an interesting theoretical point to consider. It may simply be the case that these processes are an artifact induced by the sequential structure typical of most judgment experiments. What is known, based on the results presented here, is that such cognitive processes can manifest as serial dependence in ex-

pert prognostic judgment tasks. The application of time series theory and analysis to existing data, as well as to data from future experiments, holds potential for revealing additional insights into how sequential context effects manifest in judgment tasks.

References

- Anderson, K. M., Odell, P. M., Wilson, P. W. F., & Kannel, W. B. (1991a). Cardiovascular disease risk profiles. *American Heart Journal*, *121*, 293–298.
- Anderson, K. M., Wilson, P. W. F., Odell, P. M., & Kannel, W. B. (1991b). An updated coronary risk profile: A statement for health professionals. *Circulation*, *83*, 356–362.
- Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.
- Beckstead, J. W. (2007). A note on determining the number of cues used in judgment analysis studies: The issue of type II error. *Judgment and Decision Making*, *2*, 317–325.
- Beckstead, J. W., & Stamp, K. D. (2007). Understanding how nurse practitioners estimate patients' risk for coronary heart disease: A judgment analysis. *Journal of Advanced Nursing*, *60*, 436–446.
- Box, G. E. P & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco, CA: Holden-Day.
- Brehmer, B. & Joyce, C. R. B., (1988). *Human judgment: The SJT view*. Amsterdam: North Holland Elsevier.
- Brocklebank, J. C. & Dickey, D. A. (2003). *SAS for forecasting time series* (2nd ed.). Cary, NC: SAS Institute, Inc.
- Brunswik, E. (1952). *International encyclopedia of unified science (vol I, no. 10): The conceptual framework of psychology*. Chicago: University of Chicago Press.
- Brunswik, E. (1956). Historical and thematic relations of psychology to other sciences. *Scientific Monthly*, *83*, 151–161.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah NJ: Lawrence Erlbaum Associates.
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Francisco: Academic Press.
- DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, *119*, 375–396.
- Dhami, M. K., & Ayton, P. (1998). *Legal decision making the fast and frugal way*. Poster presented at the annual meeting of the Society for Judgment and Decision Making, Dallas, Tx.
- Dhami, M. K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making*, *14*, 141–168.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 427–431.
- Durbin, J. & Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, *37*, 409–427.
- Durbin, J. & Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, *38*, 159–178.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991) Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Gregson, R. A. M. (1976). Psychophysical discontinuity and pseudosequence effects. *Acta Psychologica*, *40*, 431–451.
- Gregson, R. A. M. (1983). *Time series analysis in psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hammond, K. R., Stewart, T. R., Brehmer B., & Steinmann, D. O. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human Judgment and Decision Processes*, (pp. 271–312). New York: Academic Press.
- Harries, C. (1995). *Judgment analysis of patient management: General practitioners' policies and self-insight*. PhD dissertation, University of Plymouth.
- Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics*, *3*, 409–414.
- Johnston, J. (1984). *Econometric methods*, (3rd ed.). New York: McGraw-Hill, Inc.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME: Journal of Basic Engineering*, *3*, 35–47.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Krantz, D. H. & Tversky, A. (1971). *Conjoint measurement analysis of composition rules in psychology*. *Psychological Review*, *78*, 151–169.
- Laming, D. (1995). Screening cervical smears. *British Journal of Psychology*, *86*, 507–516.

- Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.
- Luce, R. D. & Tukey, J. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1–27.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ostrom, C. W. Jr. (1990). *Time series analysis: Regression techniques* (2nd ed.). Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–009. Newbury Park, CA: Sage.
- Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, 112, 881–911.
- Stewart, T. R. (2001). The lens model equation. In K. R. Hammond & T. R. Stewart, (Eds.), *The Essential Brunswik: Beginnings, explications, applications*, (pp. 357–362). New York: Oxford University Press.
- Tape, T. G, Kripal, J. & Wigton, R. S. (1992). Comparing methods of learning clinical prediction from case simulations. *Medical Decision Making*, 2, 213–221.
- Vlaev, I. & Chater, N. (2007). Context effects in games: Local versus global sequential effects on choice in the prisoner's dilemma game. *Judgment and Decision Making*, 2, 380–389.
- Wiener, N. (1949). *Extrapolation, interpretation and smoothing of stationary time series*. New York: Wiley.
- Wonnacott, R. J., & Wonnacott, T. H. (1979). *Econometrics*, (2nd ed.). New York: John Wiley & Sons.
- Yaffee, R. (2000). *Time series analysis and forecasting with applications of SAS and SPSS*. San Diego: Academic Press, Inc.