Effective Judgmental Forecasting in the Context of Fashion Products

Matthias Seifert, IE Business School

Enno Siemsen, Carlson School of Management, University of Minnesota

Allègre L. Hadida, University of Cambridge Judge Business School

Andreas Eisengerich, Imperial College

February 2015

Forthcoming, Journal of Operations Management

Send correspondence to Matthias Seifert, IE Business School, Maria de Molina 13, Spain (Matthias.Seifert@ie.edu); Allègre L. Hadida, University of Cambridge Judge Business School, Trumpington Street, Cambridge CB2 1AG, U.K. (a.hadida@jbs.cam.ac.uk).

Effective Judgmental Forecasting in the Context of Fashion Products

Abstract. We study the conditions that influence judgmental forecasting effectiveness when predicting demand in the context of fashion products. Human judgment is of practical importance in this setting. Our goal is to investigate what type of decision support, in particular historical and/or contextual predictors, should be provided to human forecasters to improve their ability to detect and exploit linear and nonlinear cue-criterion relationships in the task environment. Using a field experiment on new product forecasts in the music industry, our analysis reveals that when forecasters are concerned with predictive accuracy and only managerial judgments are employed, providing both types of decision support data is beneficial. However, if judgmental forecasts are combined with a statistical forecast, restricting the decision support provided to human judges to contextual anchors is beneficial. We identify two novel interactions demonstrating that the exploitation of nonlinearities is easiest for human judgment if contextual data are present but historical data are absent. Thus, if the role of human judgment is to detect these nonlinearities (and the linearities are taken care of by some statistical model with which judgments are combined), then a restriction of the decision support provided would make sense. Implications for the theory and practice of building decision support models are discussed.

Keywords: Judgmental forecasting, fashion products, lens model design, demand uncertainty, music industry, new product forecasting

Please note that an article based on portions of the same empirical data sample has appeared in Organizational Behavior and Human Decision Processes (Volume 120, January 2013). Cross-references are used where appropriate.

1. Introduction

The accurate prediction of the commercial success of newly launched products or services represents a crucial managerial problem (Steenkamp, Hofstede and Wedel 1999; Stremersch and Tellis 2004; Van den Bulte and Stremersch 2004). Generating such forecasts can be extremely difficult, particularly in environments involving fashion-oriented consumer products (hereafter referred to as "fashion products"), where the nature of the products may contain a substantial creative, artistic component, and consumer taste constantly changes (Christopher, Lowsen and Peck 2004; Hines and Bruce 2007; Hirsch 1972). In more conventional forecasting domains, for example when predicting the demand of machine spare parts in manufacturing (e.g. Sani and Kingsman 1997) or when estimating electricity demand (e.g. Taylor and Buizza 2003), large amounts of historical data are often available to calibrate decision support models and achieve high levels of model accuracy. Forecasts about the demand of fashion products, on the other hand, often lack such integral information as the demand pattern tends to be highly uncertain (Choi, Hui, Liu, Ng and Yu 2013; Green and Harrison 1973; Sichel 2009; Sun, Choi, Au and Yu 2008).

Companies producing fashion products for which consumer tastes and preferences cannot be tracked continuously often perceive themselves more as trend-setters than trend-followers (Eliashberg, Weinberg and Hui 2008). For instance, in the apparel industry, firms face the challenge of quickly commercializing new designs that are introduced during the New York or Paris Fashion Week in order to create and satisfy new consumer demand. Yet, forecasts about future sales of new clothing designs are highly volatile. They depend both on managers' ability to accurately anticipate uncertain consumer preferences and on their firm's time-to-market capability relative to its competitors. In such supply-driven environments (Moreau and Peltier 2004), conventional time series methods typically cannot be employed to predict demand with reasonable accuracy (Eliashberg and Sawhney 1994; Moe and Fader 2001; Sawhney and Eliashberg 1996). Instead, researchers have proposed several approaches to overcome the problem of model calibration when only limited and/or unreliable data are available (for an extensive review, please see Nenni, Giustiniano and Pirolo 2013). In particular, these approaches have relied on Bayesian estimation when new sales data becomes available (e.g. Green and Harrison 1973), Fourier analysis (Fumi, Pepe, Scarabotti and Schiraldi 2013), binomial distribution models (Cachon and Fisher 2000), the Croston method (Snyder 2002), two-stage dynamic sales forecasting models (Ni and Fan 2011), artificial neural networks (Au, Choi and Yu 2008; Gutierrez, Solis and Mukhopadhyay 2008; Yu, Choi and Hui 2011), fuzzy logic (Thomassey, Happiette and Castelain 2005), extreme learning machines (Sun et al. 2008; Xia, Zhang, Weng and Ye 2012), and hybrid intelligent models (Choi et al. 2013; Wong and Guo 2010). Furthermore, quick response manufacturing strategies have frequently been proposed as efficient means to shorten production lead times and gather "early sales" signaling data to reduce demand uncertainty (e.g. Cachon and Swinney 2011; Fisher and Raman 1996; Iyer and Bergen 1997). However, despite the growing variety of managerial approaches and quantitative models at our disposal, the practice of forecasting product and service success continues to crucially depend on human judgment (Sanders and Manrodt 2003, Boulaksil and Franses 2009).

Past research has therefore emphasized the importance of understanding how and when managerial judgment contributes to improving forecasting accuracy. For example, in the context of newsvendor decision making, accurately estimating product demand before the selling period is extremely important in order to minimize inventory costs and avoid lost sales (Bolton and Katok 2008; Bostian, Holt and Smith 2008; Donohue and Schultz 2006; Schweitzer and Cachon 2000). Lee and Siemsen (2015) showed that decomposing the forecasting task from the ordering task in these contexts can improve judgmental accuracy under certain conditions. Task decomposition can therefore be seen as a possible means to *debias* the so-called pull-to-center effect that has often been found in behavioral studies of the newsvendor problem (e.g. Ho, Lim and Cui 2010; Ren and Croson 2013; Su 2008). Moreover, Blattberg and Hoch (1990) demonstrated in the context of catalogue fashion sales how buyers' demand predictions could be improved by relying on an equally weighted combination of model and expert forecasts. Gaur, Kesavan, Raman and Fisher (2007) emphasized the effectiveness of using dispersion among expert judgments to predict the actual uncertainty of demand for fashion products. Eliashberg, Weinberg and Hui (2008) also concluded that managerial judgment is essential for the scheduling of motion pictures and the prediction of box office results. Furthermore, Lawrence, Goodwin, O'Connor and Önkal (2006) argued that managerial judgment is likely to prove valuable whenever the "ecological validity" of formal models is low, which is indicated by a low fit between model and forecasting environment.

The purpose of this paper is to study the conditions that determine the effectiveness of judgmental forecasting in environments involving fashion products. The sense-making mechanism underlying expert judgment is often viewed as a pattern-matching process during which forecasters perceive informational stimuli of a forecasting event and make their inference by comparing them to similar situations experienced in the past. As forecasting performance is limited by judges' ability to retrieve comparable situations from their memory, providing more *historical cases* to increase the possibility of finding a good match may therefore support their judgmental processes (Hoch and Schkade 1996). A second approach to aid decision making is to support the strengths of expert judgment by providing forecasters with better *contextual data*. Numerous studies have demonstrated that the strengths of expert judges lie in particular in their ability to diagnose new variables, recognize abnormal "broken leg cues", and evaluate information that is otherwise difficult to quantify (Blattberg and Hoch 1990; Einhorn 1974). Providing judges with more contextual data may therefore enable them to make better sense of a forecasting event. Such contextual data might comprise product-specific information relative to promotional activities, manufacturing data, and more general domain knowledge including competitor data or macroeconomic forecasts (Lawrence et al. 2006).

We extend Hoch and Schkade's (1996) work by offering empirical insights into the usefulness of decision support approaches relying on historical cases and on contextual data in a setting characterized by high uncertainty. Whereas research on the general accuracy and appropriateness of judgmental forecasts has a long-standing tradition, only few studies have attempted to decompose the forecasting context into different types of knowledge components in order to examine their effect on judgmental performance. Among these, Blattberg and Hoch (1990), Stewart, Roebber and Bosart (1997), and Seifert and Hadida (2013) have studied whether judgmental forecasts can add value beyond the predictions of linear models. However, these studies rest on the assumption that human judgment is capable of approximating the linear regression model of the environment fairly well. This assumption appears questionable, at least in some forecasting contexts, when considering that the information processing capacity of forecasters is limited. In addition, while Lee and Siemsen (2015) as well as Seifert and Hadida (2013) acknowledge that differences in forecasting effectiveness may depend on task structure, a more systematic decomposition of the forecasting environment is required to fully understand how task characteristics fundamentally influence judgmental performance and how decision support systems should be designed to improve predictive accuracy.

To study how historical demand anchors and contextual anchors interactively influence the performance of *human judgment*, we employ a judgment analysis approach (Cooksey 1996; Hammond 1996). Judgment analysis allows us to analyze managerial predictions beyond forecasting accuracy by decomposing judgments into two different components. First, we focus on the degree to which a manager's interpretation of a forecasting event matches the efficiency of a linear model. Second, we measure the extent to which a managerial judgment can reduce the residual variance of the linear model by interpreting contextual knowledge surrounding a forecasting event and, hence, add predictive value over and beyond the linear model. Our empirical context is the music industry, which, due to its creative and artistic nature, can be understood as a typical market for fashion products (Santagata 2004). Specifically, we study forecasts about the Top 100 chart entry positions of upcoming pop music singles. While past research has primarily focused on sales predictions in the domain of fashion retailing, the music sector appears to be a particularly interesting forecasting domain where product success is highly contingent on the subjective evaluation of a multitude of independent industry gatekeepers (Vogel 2007).[1] Our study reveals that when the primary concern is to maximize predictive accuracy and forecasts are based on human judgment *only*, providing both types of decision support data is likely to minimize forecast errors. However, more interestingly, our results also showcase the ambivalent nature of decision support anchors as the presence of both types of data appears to improve forecasters' ability to exploit nonlinearities, while impairing their effectiveness of interpreting linearities in the task environment. In fact, our field data indicate that the exploitation of nonlinearities is easiest for human judgment if contextual data are present but historical data are absent. Thus, if the objective is to use human judgment to maximize the explanation of nonlinearities surrounding the forecasting task, we suggest that decision support should be restricted to contextual data anchors and subsequently combined with the prediction of a statistical linear model to optimize combined forecasting performance. The next section introduces the theoretical background of our research. Section 3 provides an overview of the data collection context and methods. In sections 4 and 5, we report the model results, and discuss them in light of existing literature. Section 6 concludes the study, discusses its limitations and offers directions for future research.

2. Theoretical Background

2.1. Judgment Analysis

Judgment analysis is rooted in the assumption that forecasting environments can be described in terms of the probabilistic relationships between relevant informational cues (i.e., the predictor variables) and an ecological criterion (i.e., the forecasting event) (e.g., Dougherty and Thomas 2012; Einhorn 1970, 1971; Hammond and Summers 1972). Past research has frequently tested a

[1] In the music industry, rank predictions are highly correlated with actual sales forecasts. Specifically, our secondary data analysis of Top 40 *album* chart ranks in the UK between 1996 and 2003 and their associated average weekly sales levels reveal a strong negative correlation of $r = -0.74$, indicating that smaller ranks are associated with higher sales levels. A corresponding analysis of the Top 40 *singles* charts resulted in a similarly strong negative correlation of $r = -0.68$.

forecaster's ability to exploit linear relations between predictor variables and forecasting event (hereafter called "linearities") by analyzing how well a linear regression model of the environment fits subjects' judgments (e.g., Dhami and Harries 2001; Karelaia and Hogarth 2008). Because a linear model describes the task fairly well in *stable* environments, judgment analysis studies have often assumed any type of nonlinear relation between predictor variables and forecasting event (hereafter referred to as "nonlinearities") to be negligible random noise in their research designs (e.g., Dunwoody, Haarbauer, Mahan, Marino and Tang 2000; Hammond and Summers 1972). However, in many organizational settings, reducing the forecasting context to linear cue-criterion relations may produce an overly simplistic representation of the task environment (Seifert and Hadida 2013). When forecasting the demand for an upcoming novel in the publishing industry, a linear model would fail to exploit important nonlinear relationships, for instance between demand and the book's story line, genre and/or author information. Yet, such informational cues may prove extremely useful for reducing the residual variance of a linear model as they could help the forecaster to understand, for instance, whether the book satisfies the demand arising from a recent hype surrounding Scandinavian crime stories.

In this article, we regard the residual variance of linear models as an important source of information that judges may exploit by relying on their familiarity with the task and their domain-specific expertise (Lawrence et al. 2006). Similarly, in a study of temperature and precipitation forecasts, judgment analysis was employed to demonstrate how systematically exploiting nonlinear relations between predictor variables and forecasting event could lead to improved forecasting accuracy (Stewart et al. 1997). The same study also revealed a relationship between the predictive value of nonlinearities and the type of information available to the forecaster. In particular, when the informational cues available to forecasters were highly complex, the study not only demonstrated the limited usefulness of linear model forecasts, but also the increasing importance of nonlinearities for improving forecasting accuracy. We next discuss the extant literature on forecasting fashion products and then turn to the relationship between task properties and forecasting accuracy.

2.2. Forecasting the Demand of Fashion Products

Fashion products are sometimes referred to as *semaphoric* goods due to their creative and artistic nature and to the subjective, symbolic value associated with them (Santagata 2004). Markets for fashion goods can generally be described in terms of four distinctive features (Christopher et al. 2004; Nenni, Giustiniano and Pirolo 2013): (1) short (seasonal) product life-cycles, which can usually be measured in months or weeks; (2) high volatility due to instable, nonlinear product demand; (3) low week-by-week and/or item-by-item predictability; and (4) highly impulsive purchasing decisions of consumers, which crucially require products to be available in the store. Given these characteristics, firms operating in fashion markets need to maintain a high level of flexibility during the manufacturing process and very short lead times (Choi et al. 2013). Consequently, conventional organizational structures and decision support approaches often prove inadequate (Christopher et al. 2004).

The music industry matches the characteristics of a market for fashion products. Similar to firms in the apparel, publishing and movie industries, music companies continuously face the challenge of creating and satisfying consumer demand for new *fads and fashions* in order to secure organizational survival (Hines and Bruce 2007; Hirsch 1972; Peterson and Berger 1971). Typically, music firms do not only forecast ordinary sales targets as the basis for their production planning and marketing strategies. They also generate predictions about the entry positions of new music releases in the Top 100 charts. Chart entry predictions are extremely complex, because record companies can only promote the release of a single to a limited extent. In particular, promotion depends on several independent gatekeepers such as radio and tv stations, press, retailers, and online communities. These gatekeepers rely on their own selection criteria, processes and time lines to decide to feature, review and/or include a product in playlists prior to its official release date (Hadida and Paris 2014; Vogel 2007). In addition, chart predictions are further complicated by the fact that record companies are required to publicly announce all upcoming single launches several months in advance. Release schedules may consequently partly reflect a company's strategy to avoid or capitalize on competing product launches that were known at the time of scheduling. The chart entry position itself represents a direct measure of sales performance, since the overall 100 best-selling singles in any given week are included in the charts of the following week.[2] Because sales typically peak during the release week, the chart performance of a single tends to be highest at the time of chart entry and then declines during the subsequent weeks until the single drops out of the top 100-selling products.[3] In general, separate music charts exist for both single song releases in a given week and for full album releases. The chart position of a music single may hereby serve as an early sales signal for an upcoming album release, as it allows companies to test the market response and gather information for estimating the demand of a full album. Singles therefore play an important role in the quick response inventory management strategies of music firms (Choi, Li and Yan 2004; Serel 2009). Taking this into account, although the advent of digital distribution has led to a drastic decrease in the cost of forecast errors, the music industry context has the advantage of providing very frequent introductions of fashion products with publicly observable market outcomes. This advantage is important for our study.

2.3. Hypotheses

In this paper, we study forecasts in the domain of fashion products by focusing on how the presence of two types of data, *historical data* and *contextual data*, supports a judge's ability to process linearities and nonlinearities in the forecasting environment. Although each of the two types of data individually has received considerable attention in the judgmental forecasting

[2] Since the chart entry of a new single is always relative to the sales performance of other, previously launched singles, it does not represent an independent event. However, our interviewees noted that the influence that other singles included in the charts exert is not well understood because their *actual* sales level is not observable at the time of the forecast. For modeling purposes, we therefore treat the interaction between singles in the charts as random noise.

[3] For further details about the industry context please see Seifert and Hadida (2013), in which part of the same dataset has been used to analyze the performance of forecast aggregations.

literature (e.g., Gaur et al. 2007; Hoch and Schkade 1996; Lawrence et al. 2006), to the best of our knowledge, none of the existing studies has systematically examined their joint effect on a manager's judgmental forecasting performance.

Historical data. Prior research has established that historical demand data can reduce judgmental forecasting uncertainty (Gaur et al. 2007; Lawrence et al. 2006). Consequently, historical information has been argued to increase the transparency of means-end relationships (Fellner 1961; Frisch and Baron 1988) by shedding more light on (a) the probability distributions that link predictive variables and forecasting event (Camerer and Weber 1992; Wood 1988); (b) the transparency of path sequences; and/or (c) the appropriateness of organizing principles, which could be utilized to assess the relative importance of informational cues and, hence, systematically exploit them (Hamm 1987; Hammond 1996; Steinmann 1976; Wood 1988). In time series forecasting, however, where historical information mostly relates to the length of a time series, research findings have been somewhat contradictory: Although there is some evidence indicating forecasters' proficiency in using historical data for detecting trends, seasonality, randomness and discontinuities (Lawrence et al. 2006; Lawrence and Makridakis 1989), others have argued that quite minor changes in a series and in the presentation of a task can substantially impair judges' forecasting ability (Andreassen and Kraus 1990; Goodwin and Wright 1993). Behavioral experiments based on the newsvendor problem have shown that managerial judgments can be severely influenced by the recent history of a time series (Bostian et al. 2008; Schweitzer and Cachon 2000). Furthermore, Kremer, Moritz and Siemsen (2011) observed that in contexts of stable time series, forecasters have a tendency to engage in "demand chasing": They predict the next step in a time series according to simple error-response learning, despite having enough data available to determine that errors are mostly driven by noise in the time series.

Taking this into account, research has directly related judgmental forecasting performance to the level of noise in the underlying data series (Harvey 1995; Harvey, Ewart and

West 1997). In particular, in uncertain environments forecasters appear to reproduce noise levels in their judgments. This can be understood as a reflection of their inability to disentangle useful informational signals from random variability. The observation that human judges do not perform well as "intuitive statisticians" in noisy environments is not new (Brehmer 1978; Harvey et al. 1997; Kahneman and Tversky 1973). In fact, when regressing forecasters' judgments onto informational cues, some residual variance typically remains unexplained (Harvey et al. 1997). In his study of multiple-cue probability learning tasks, Brehmer (1978) examined whether the residual variance component of forecasters' judgments contained systematic deviations from linearity given different levels of data noise. His study showed that judgmental inconsistencies were unsystematic and therefore could not be explained by forecasters' use of nonlinear functional rules in any of the tested settings.

In light of this discussion, when forecasting the commercial performance of fashion products, the level of noise underlying historical data is likely to have implications for forecasters' ability to interpret linearities as well as nonlinearities in the environment. Specifically, Brehmer's (1978) study indicates that noise appears to deteriorate judgmental performance in general and, hence, is likely to affect the effectiveness of both information processing mechanisms. Thus, data noise appears to create uncertainty about how to systematically weight informational cues in the task environment. Furthermore, data noise decreases the reliability of human judgments, because it creates ambiguity regarding the *type* of linear or nonlinear function that could best describe the relationship between predictor variables and forecasting event (Einhorn 1971; Harvey 1995; Hogarth 1987). Research in another market for fashion products, cinema, also found no significant impact of past historical demand anchors such as the prior box office performance of leading actors (Zuckerman and Kim 2003) or directors and producers (Elberse and Eliashberg 2003; Zuckerman and Kim 2003) on a film's box office receipts. We therefore expect that the provision of historical demand anchors neither improves judges' ability to better utilize linear models, nor does it enhance their ability to process nonlinearities among predictor variables in a more effective manner.

HYPOTHESIS H$_1$: *When forecasting the commercial performance of fashion products, the presence of historical demand data neither improves forecasters' ability to process linearities (i.e. by approximating the accuracy of a best fitting linear model) nor does it improve forecasters' ability to process nonlinearities in the task environment.*

Contextual data. Contextual data surrounding a prediction task can serve as an important anchoring point, which is then up- or down-adjusted by the forecaster to arrive at a judgment (Lawrence et al. 2006). These data refer to all non-historical information on a forecasting event, including predictions derived from market research, past and future promotional plans, and competitor, manufacturing and macroeconomic data. Such non-historical information can often have a more subjective, qualitative nature and enable judges to obtain a more comprehensive understanding of the *nonlinear* relationships that describe a forecasting event (Blattberg and Hoch 1990; Seifert and Hadida 2013). Contextual data therefore allows forecasters to develop a more holistic mental representation of the prediction task; and may ultimately result in higher predictive accuracy. In addition, prior research has shown that when contextual data are available, cognitive shortcuts, such as pattern matching or the use of an equal weighting procedure for assessing informational cues, are frequently adopted to efficiently integrate task information (Clemen 1989; Einhorn and Hogarth 1975; Hoch and Schkade 1996). Therefore, the presence of contextual data is likely to improve judgmental accuracy by allowing forecasters to better exploit nonlinearities.

However, because informational cues in highly uncertain environments can represent imperfect signals of a real-world event (Olshavsky 1979; Payne 1976; Slovic and Lichtenstein 1971), contextual anchors may also lead to systematic errors if they entice forecasters to see patterns where none exist. In particular, previous studies have shown that people frequently fail to judge randomly generated patterns in a sequence of items as originating from true randomness

(Harvey 1995). Instead, they adjust their forecasting judgments in an attempt to make them become more representative of the underlying data structure. In fact, the representativeness heuristic has often served as an explanation for why people fail to make regressive predictions (Harvey et al. 1997; Kahneman and Tversky 1973). Contextual anchors thus have an ambivalent nature: although they enrich forecasters' understanding of the environment, they are also likely to increase the complexity of a forecasting task. This happens particularly through (a) an increase in the number of informational cues available for making a judgment, (b) a decrease in cue inter-correlations, (c) the necessity of applying nonlinear functions and unequal cue weightings, and (d) the absence of a simple organizing principle for integrating information (Hamm 1987; Steinmann 1976). An increase in the amount of contextual data is therefore likely to come primarily at the expense of forecasters' ability to process linearities in the environment. Specifically, previous research has demonstrated that the consistency of exploiting linear relationships between predictor variables and forecasting event decreases as the complexity of a forecasting task increases (Einhorn 1971; Lee and Yates 1992). In sum, we expect contextual anchors to have a positive effect on forecasters' ability to process nonlinearities and a negative effect on their ability to process linearities in the task environment.

HYPOTHESIS H_2: *When forecasting the commercial performance of fashion products, the presence of contextual anchors is negatively associated with forecasters' ability to process linearities, and positively associated with their ability to process nonlinearities in the task environment.*

Joint effects of historical demand and contextual anchors. We now turn to how historical demand anchors and contextual anchors may jointly affect a judge's ability to process linearities and nonlinearities in the task environment. Hypothesis H_2 states that when forecasting the commercial performance of fashion products, contextual data emit signals that are primarily useful for processing nonlinearities. As per Hypothesis H_1, historical demand data, on the other hand, are unlikely to improve forecasters' ability to exploit linear or nonlinear relationships between predictor variables and forecasting events. If no historical demand data are available, it seems likely that the impact of contextual anchors on processing linearities will be smaller than if both historical and contextual anchors are provided. In the joint presence of the two types of data, forecasters may be able to search for (misleading) confirmatory evidence between signals. Thus, in the case of linearities, the negative effect of contextual data anchors may be amplified in the presence of historical data. Similarly, the two anchors also interactively influence forecasters' efficiency in processing nonlinearities in the task environment. Efficiently exploiting nonlinearities is relatively more difficult than exploiting linearities, because there may be many different types of nonlinear functions that could potentially describe the forecasting data. Stewart et al. (1997) and Seifert and Hadida (2013) have shown that judges are likely to be most proficient in exploiting nonlinearities when the predictability of the forecasting environment is low. The presence of both types of anchors, however, is likely to lower forecasters' ability to extract nonlinearities, because the historical demand anchor may provide (false) unrepresentative clues about forecasters' accurate perception of how nonlinearities should be evaluated. In the absence of historical demand data, we thus expect the positive impact of contextual anchors on processing nonlinearities to be greater.

HYPOTHESIS H_{3a}: *In the presence of historical demand data, contextual data will have a greater negative effect on forecasters' ability to exploit linearities than in the absence of historical demand data.*

HYPOTHESIS H_{3b}: *In the presence of historical demand data, contextual data will have a smaller positive effect on forecasters' ability to exploit nonlinearities than in the absence of historical demand data.*

3. Methods

Our approach is based on a Brunswikian lens model (Brunswik 1956) that is frequently used in judgment analysis. We develop two distinct explanatory variables to assess a judge's ability to integrate linear and nonlinear forecasting information. Specifically, our dependent variables represent effectiveness measures that capture the degree to which judgmental forecasts deviate from the predictions of the best fitting linear model between the available data (cues) and the product success variable the data should predict (criterion). Our first dependent variable indicates a judge's effectiveness in analyzing linear cue-criterion relations. The variable focuses on the extent to which the linear model of the expert judge approximates the best-fitting linear model of the forecasting environment. Our second dependent variable measures a judge's ability to process nonlinearities. The variable refers to the amount of variance that is unaccounted for by the linear model of the environment, and which instead can be explained by respondents' individual judgments. We further derive two independent sets of predictor variables from the empirical context to operationalize historical data and contextual data. A multivariate analysis of variance is used to conduct our hypotheses tests.

"Figure 1 here"

An overview of the full judgment analysis method is provided in Figure 1. In particular, the models are constructed around the ecological criterion (*O*), which represents a probabilistic function of a number of observable environmental cues (X_i) (Cooksey 1996). These informational cues may be redundant, and a full model of the ecological criterion may be specified as follows:

(a) $$O = \widehat{M}(X_1, X_2 \dots X_n) + \varepsilon,$$

where $\widehat{M}$ denotes the linear, best-fitting model between cues and the criterion.[4] The model can be interpreted as a benchmark to which linearities in the task environment can be integrated.[5] The error ε represents the residual, unmodeled variance; it comprises random error and

[4] While ordered logistic regression is typically used when the task involves the prediction of rank orders, it would prove inefficient in the context of our forecasting task. Specifically, since our dependent variable contains 100 levels, the resulting model would require an estimation of 100-1 equations, which would make a meaningful analysis become extremely challenging. Moreover, such a model would impose major restrictions on the characteristics of the data sample needed for making meaningful inferences. Instead, scholars have argued that linear regression models can approximate the predictions of logit models very well if few outliers exist (e.g. Iman and Conover 1979; Lattin, Carroll and Green 2003). For this reason, we only included new singles that actually entered the Top 100 singles charts, which led to an exclusion of 9 observations from the sample. The relatively small number of outliers can be partially explained by the fact that music singles generally enter the charts at their peak position (followed by a decline in the subsequent weeks) and by companies' scheduling strategies for upcoming single releases, which aim at avoiding cannibalization effects with other singles released during the same week.

[5] In the context of our study, we only included linear task relations in our ecological models in order to clearly distinguish between judges' ability to process linear and nonlinear cue-criterion relations. Future studies could also extend our models to include nonlinearities as a more conservative threshold for judges' ability to pick up unmodeled residual variance.

nonlinear relations between the informational cues and the criterion (Blattberg and Hoch 1990). Our participants observe the predictor variables before they make a forecasting judgment (*Y*) regarding *O*. When regressing the environmental cues onto *Y*, we can derive a second linear model $\widehat{M}^*$, which captures the extent to which forecasting judgments can be explained as a function of the informational cues given to our participants. Therefore, a participant's judgment regarding the ecological criterion can be modeled as follows:

(b) $$Y = \widehat{M}^*(X_1, X_2 \dots X_n) + \varepsilon^*$$

where the error term ε^* represents the residual variance unexplained by the judge's model $\widehat{M}^*$. Given that judges rely on the same informational cues to generate individual predictions (*Y*) as the best-fitting linear model, the following lens model equation ($R_{Y,O}$) can be used to establish the relationship between judge and task environment (Stewart et al. 1997; Stewart 2001; Tucker 1964):

(c) $$R_{Y,O} = R_{O,X} G R_{Y,X} + U\sqrt{1 - R_{O,X}^2}\sqrt{1 - R_{Y,X}^2}.$$

In equation (c), $R_{O,X}$ denotes the correlation between target event *O* and the best-fitting linear model of the environment $\widehat{M}$, whereas $R_{Y,X}$ measures the correlation between the judge's model $\widehat{M}^*$ and judgments *Y*. Our two focal points of analysis, however, are *G* and *U*. Coefficient *G* is often called a matching index (Cooksey 1996; Tucker 1964) because it describes judges' effectiveness in processing linear cue-criterion relations. More specifically, we use G^2 as a measure for the amount of task variance that is jointly explained by $\widehat{M}$ and $\widehat{M}^*$. Moreover, *U* represents a semipartial correlation coefficient that indicates the strength of association between the unmodeled components of a target event and judgmental forecasts ($r_{\varepsilon,\varepsilon^*}$). We interpret *U* as an important proxy of judges' skill in analyzing nonlinearities in the forecasting environment. We note, however, that *U* may be somewhat distorted by random noise as well as omitted variables that are neither captured by the linear model of the judge nor by the best-fitting linear model. When squared, U^2 includes the unique contribution of judges' forecasts *Y* in explaining the unmodeled variance in the task environment (Cohen and Cohen 1975). Although our interpretation of the residual component U^2 cannot be used to precisely isolate forecasters' skill from the deficiencies of the linear model, several studies have shown that it does represent a useful indicator for the validity of judges' ability to analyze nonlinearities (e.g., Blattberg and Hoch 1990; Seifert and Hadida 2013). Together, G^2 and U^2 measure the effectiveness of judgments by decomposing them into forecasters' linear and nonlinear achievements.

4. Empirical Setting

Our field data consist of judgments about new product releases in the music industry. More specifically, we analyze judgmental predictions about the entry positions of new, previously unreleased pop music singles on the national Top 100 singles sales charts. We began our data collection by conducting 23 semi-structured interviews with senior managers at the "big 4" major record companies, which together account for more than 80% of the global music market share (IFPI 2014). We collected our interview data in the two largest domestic markets within the European Union: Germany and the United Kingdom. Our interview participants can be regarded as experts in the industry, given the "up-or-out" hiring structure in major record companies, which ensures that only managers with the most successful track records of placing "hits" are employed (Vogel 2007). The average industry tenure of our interviewees was 10 years, and all held divisional or regional responsibilities. The main purpose of the interviews was to elicit potential predictor variables surrounding our forecasting event, which would enable us to specify the best-fitting linear model of the environment. We used an iterative process, during which interviewees were given the opportunity to revise their initial list of predictor variables to reduce the possibility of omitting relevant variables in the subsequent model building process. This revision strengthened the validity of our measure for forecasters' ability to analyze nonlinearities in the environment (U^2).

The second phase of our data collection involved quantitative predictions made by 92 Artist & Repertoire (A&R) managers from major and medium-sized record companies:

approximately one-half in Germany, and the remaining in the United Kingdom. All managers were contacted through the leading international A&R managers' association "HitQuarters". Participants included in the sample had placed 21.74 hits on average on previous Top 100 charts, and had a mean industry experience of 7 years.

Participants generated forecasting judgments by completing four online questionnaires over a period of 12 weeks. The questionnaires were based on the predictor variables identified during the initial interviews and contained the profiles of a set of yet unreleased pop music singles.[6] Participants were asked to estimate the imminent chart entry position of the singles, and to indicate their confidence in the judgment they provided.[7] The cue profiles were generated by drawing on information obtained from marketing research firms, chart compilation companies, retailers, key media firms, record companies, and the Internet.

"Table 1 here"

Our data collection period took place in four batches and allowed for a period of two to three weeks between prediction and entry in the music charts, so that every pop-music single was exposed to the media for an equal duration. Altogether, we generated a sample of 210 prediction cases, and each participant provided 40 forecasts about the pop music singles. Because participants made judgments about real (rather than artificially generated) prediction cases, it was not possible to assign singles in a truly randomized manner. Instead, singles included in each questionnaire were dictated by the record labels' release schedules in a particular week. We created four batches of judges (two for each country sample); and alternated their participation in two-week intervals. With each round of participation, subjects provided up to 10 forecasts about the upcoming releases in a particular week. They had one week to return the questionnaires, and response dates were recorded to control for any time-based effects on the accuracy of participants' predictions.

[6] The appendix provides two examples of the cue profiles used.
[7] We excluded the analysis of participants' judgmental confidence from the scope of this paper. However, the data are available from the authors upon request.

Historical demand and contextual data. It became clear from our initial interviews that the task of forecasting music singles' chart entry positions may depend on two important factors: (1) the intensity and success of the promotional efforts conducted one month ahead of the product release and (2) the distinction between established artists and new, unknown artists.

We use factor (1) as a proxy for contextual anchors. They include the following informational cues: *the peak position of the single in the radio airplay charts*; *the production of a music video and its performance in the music video charts; critical reviews of the upcoming single in the most influential industry magazines*; and *the amount of money spent by the record label to finance general retail campaigns*. We also considered whether the single formed part of a *previously released album*, which may provide additional information about the single's potential chart success. Even so, each media outlet employs its own criteria and institutional procedure for selecting and ranking music singles, and signals regarding a single's success potential are typically imperfect and conflicting (e.g., achieving a high ranking in the radio airplay charts and, at the same time, a low rating in the music press).

We use factor (2) as a proxy for the presence of historical demand anchors. New artists lack an existing track record of previous successes or an established fan base, both of which could be valuable indicators of future success. As a consequence, they decrease the predictability of the environment. We included the following historical chart success variables in our study: *the upcoming single's inclusion in the US billboard charts as a key reference market*; and *the mean chart-entry positions of previously released singles and albums in domestic and foreign markets*. In addition to the two identified categories of environmental cues, we also collected general information about the single itself, such as: a *picture of the artist*; *information about the producer and record label, the size of the artists' online fan base* and a 30-second *audio sample*.

5. Results

Consistent with previous research decomposing forecasting task environments (Dunwoody et al. 2000), Table 2 provides an overview of the descriptive statistics associated with our four decision support conditions. The table includes information about (1) the number of cues available for making a judgment, (2) the degree of redundancy among the cues (measured as the mean inter-correlation (r_c) among all relevant cues), and (3) the degree to which the environmental cues are equally weighted (measured by the standard deviation of β weights derived from the environmental model). Inter-correlation between predictors is highest when both contextual and historical data anchors are present, whereas all other decision support conditions exhibit a negligible strength of association between predictor variables. In addition, the two conditions in which no historical data anchor is available are associated with the greatest standard deviations of regression weights.

"Table 2 here"

To build linear best-fitting models of our task environment, we first compiled a broad list of informational cues that "domain experts" perceived as relevant for generating predictions of pop music singles' chart success. Next, we tested a series of candidate models, using stepwise selection to investigate the regression fit of all possible combinations of variables (Mentzer and Moon 2005). The resulting models contain 7 predictor variables in the absence of both contextual and historical anchors, 8 variables when only historical information is available, 9 variables when only contextual data are present and 12 variables when both types of decision support are available. Because the participants did not all generate forecasts regarding the identical subsamples of cue profiles, we also adjusted the resulting models for model shrinkage. Table 3 provides an overview of the cross-validated results. The results indicate a relatively good model fit, despite average model shrinkage of 13%.[8] No significant difference was observed between the two country samples. In addition, our models indicate an unequal, compensatory distribution of cue weights, in that a low score on some of the higher weighted cues (e.g., airplay chart position or previous number of hits) could be compensated for by a high score on some of

[8] We performed cross-validation analyses for all linear models and judgments obtained from the participants to control for model shrinkage (Blattberg and Hoch 1990). In the case of the statistical model, 10 random samples were drawn in which one-half of the dataset was used to simulate predictions about the remaining data. The participants were tested pair-wise using a double cross-validation method, and averaged to represent their associated participant category (Cooksey 1996).

the less important cues (e.g., whether an album had been released at the same time as the single) (Karelaia and Hogarth 2008).[9]

"Table 3 here"

We constructed separate linear models for the environment ($\widehat{M}$) and each of the judges ($\widehat{M}^*$), both of which show a higher fit when historical data are available ($\Delta R^2_{O,X}$ = 0.26; $\Delta\left(GR_{O,X}\right)^2$ = 0.23). We then tested whether the lower model fit in absence of historical data could simply be explained by the smaller number of predictor variables in the model or merely by the fact that the artist is "new". For this purpose, we re-modeled forecasting judgments about established artists, but this time without including the historical demand information. Although the absence of such data does indeed substantially reduce model fit ($C^+/H^{(+)}$: $1-R^2$=0.55 and $C^-/H^{(+)}$: $1-R^2$=0.63), judgments about brand new artists still remain less predictable in comparison with those about established artists. Hence, although historical demand data explain a large part of the difference in predictability between established and new artists, they don't fully account for it.

As a next step, we computed the mean values of the judges' ability to process linear (G^2) and nonlinear (U^2) cue-criterion relations in each of the four decision support conditions and for all forecasters in our sample. The linear model of both environment and judges reveals the highest forecasting accuracy when both historical demand and contextual anchors are present (C^+/H^+: $R_{O,X}$ = 0.82; $R_{Y,O}$ = 0.78), followed by the condition in which only historical demand anchors are available (C^-/H^+: $R_{O,X}$ = 0.71; $R_{Y,O}$ = 0.72), the case in which only contextual anchors are present (C^+/H^-: $R_{O,X}$ = 0.62; $R_{Y,O}$ = 0.68), and the condition in which neither anchor is provided (C^-/H^-: $R_{O,X}$ = 0.53; $R_{Y,O}$ = 0.57). Because our hypotheses rest on the

9 To ensure that our final models did not omit potentially important predictor variables, we also compared the variables included in these final models to the relative importance ratings of the cues that we elicited while interviewing senior executives in the industry. When variables were highly redundant and when their individual inclusion in the models did not significantly change the total variance explained, we selected those variables that received the highest subjective importance rating.

premise that the forecasting event can be at least partially characterized in terms of nonlinear relationships between environmental cues and chart positions, we conducted a series of Ramsey reset tests (Ramsey 1969) to investigate whether such nonlinearities exist in the forecasting residuals. In particular, we used power functions to create new predictor variables from the linear model predictions and added them stepwise to our models. The results show significant *F*-tests for both forecasting conditions in which historical data anchors are absent (C^{-}/H^{-}: $F = 5.834$, $p < 0.05$; C^{+}/H^{-}: $F = 10.981$, $p < 0.01$). Moreover, when historical demand anchors are present, we only find a significant test result when no contextual anchor is available (C^{-}/H^{+}: $F = 6.619$, $p < 0.05$); but not when both types of anchors are provided to judges (C^{+}/H^{+}: $F = 0.309$, $p =$ n.s.). Hence, our analysis indicates that the unexplained model variance contains both random error *and* nonlinearities in three out of the four conditions under investigation. In addition, we also created a plot of the correlation coefficients between *G* and *U* in each of the four decision support conditions (Figure 2). The plot shows that the correlation between linear and nonlinear judgmental achievement is strongest when only contextual decision support, but no historical data are available (C^{+}/H^{-}; $R_{GU} = 0.39$). This is followed by the decision support condition, in which neither of the two types of data anchors are present (C^{-}/H^{-}; $R_{GU} = 0.32$). Finally, linear and nonlinear forecasting components show the lowest level of redundancy when either both types of decision support are available (C^{+}/H^{+}; $R_{GU} = 0.26$) or when only historical data anchors are provided to the judge (C^{+}/H^{-}; $R_{GU} = 0.24$).

"Figure 2 here"

We tested our hypotheses by conducting a multivariate analysis of variance, in which contextual anchors and historical demand anchors serve as predictors of forecasters' ability to process linear and nonlinear cue-criterion relations. MANOVA was used to enable us to detect any kind of systematic relationship between the two dependent variables under investigation, which would otherwise be unobservable when relying on separate univariate analyses of variance.

The MANOVA results reveal significant Pillai's Trace coefficients for both independent variables (for contextual anchors: $T = 0.56$, $p < 0.01$, and for historical demand anchors: $T = 0.68$, $p < 0.01$). Moreover, the multivariate analysis also shows a significant interaction term with a Pillai's Trace value of $T = 0.56$ and $p < 0.01$. Based on these findings, we then tested our hypotheses by conducting six significance tests at a *p*-level of $p < 0.008$ to control for potentially inflated Type I errors.

Hypothesis H_1 postulates that the provision of historical demand anchors neither improves forecasters' ability to process linearities nor their ability to interpret nonlinearities in the task environment. Our MANOVA results provide support for this hypothesis in the following way: First, the *absence* of historical demand data leads to a slight improvement in exploiting linearities ($\Delta G^2{=}0.03$, $F = 22.379$, $p < 0.001$, $\eta^2 = 0.15$). We further explored the statistical power of this effect by utilizing the G*Power program (Faul, Erdfelder, Lang and Buchner 2007). In particular, based on the F-tests underlying our MANOVA analysis, we used *Cohen's* f^2 as our baseline effect measure as well as an alpha level of $\alpha = 0.01$ and conducted a series of post hoc tests in order to determine the odds (1- β) that the identified effect was indeed present. While a coefficient value of $f^2 = 0.01$ indicates a small effect size, coefficient values of $f^2 = 0.06$ and $f^2 = 0.16$ are generally considered to represent moderate and large effect sizes, respectively (Cohen 2013). In the case of our music data, the results of the conducted post hoc tests show a large *Cohen's* f^2 coefficient of $f^2 = 0.18$ with a probability of $(1 - \beta) = 0.99$ that the effect was not resulting from a Type II error.

Second, the data reveal that in the absence of historical demand data forecasters also appear to be more effective in exploiting nonlinearities ($\Delta U^2 = 0.06$, $F = 216.726$, $p < 0.001$, $\eta^2 = 0.64$). This finding is strongly supported by our post hoc power analyses ($f^2 = 1.78; (1 - \beta) = 1.0$). In particular, when historical data are present, judges are only capable of picking up

7% of the unexplained variance in the best-fitting linear model, compared to 14% if no such anchors are available.

We then tested the impact of contextual anchors (Hypothesis H_2), which we predict to be negative when processing linearities and positive when interpreting nonlinearities in the forecasting environment. Our MANOVA results indicate a highly significant, negative main effect for the relationship between the presence of contextual anchors and processing linearities ($\Delta G^2 = -0.06$, $F = 130.138$, $p < 0.001$, $\eta^2 = 0.51$), which is also associated with a large post hoc effect size ($f^2 = 1.04;\ (1-\beta) = 1.0$). This finding implies that judges are less efficient in analyzing task linearities when contextual information is available. Thus, in the absence of such anchors, forecasters' performance is more likely to approximate the accuracy of the best-fitting linear model of the environment. When considering the effectiveness of processing nonlinearities, our data show a significant, positive main effect ($\Delta U^2 = 0.03$, $F = 39.973$, $p < 0.001$, $\eta^2 = 0.24$), which indicates that the presence of contextual anchors facilitates forecasters' interpretation of nonlinearities. Our power analyses indicate a moderate effect size associated with this relationship ($f^2 = 0.32; (1-\beta) = 1.0$), which lends support to Hypothesis H_2.

Hypotheses H_{3a} and H_{3b} test the interactive effects of both contextual and historical demand data on the effectiveness of processing linear and nonlinear cue-criterion relations. Hypothesis H_{3a} posits that the presence of contextual data influences forecasters' ability to analyze linearities to a greater (negative) extent when historical demand data are available. Conversely, Hypothesis H_{3b} predicts that in the presence of historical demand data, contextual data have a smaller (positive) effect on forecasters' ability to process nonlinearities. To test both hypotheses, we analyzed the significance levels of each interaction term obtained in the MANOVA model and plotted two interaction graphs, with historical demand data on the x-axis and contextual data illustrated by separate lines (Figure 3).

"Figure 3 here"

We observe significant interaction effects for both dependent variables (G^2: $F = 141.593$, $p < 0.001$, $\eta^2 = 0.53$; U^2: $F = 9.141$, $p < 0.003$, $\eta^2 = 0.07$). Specifically, in the absence of historical data, the first interaction graph shows that contextual data have a negligible impact on a judge's effectiveness in exploiting linear cue-criterion relations, whereas in the presence of historical data, providing contextual data leads to a difference in G^2 of almost 12%. The statistical power of this finding is strong as our post hoc tests show considerably large effect size measures ($f^2 = 1.13; (1 - \beta) = 1.0$). Therefore, the first interaction graph validates the relationship proposed in Hypothesis H_{3a}. Consistent with H_{3b}, our analysis indicates that the unavailability of historical demand data significantly amplifies the effect of contextual data on forecasters' ability to process nonlinearities ($\Delta U^2 = 0.04$ and $\Delta U^2 = 0.01$, respectively), which is supported by observing moderate effect sizes in our post hoc examination ($f^2 = 0.08; (1 - \beta) = 0.96$). In other words, historical data moderates the effect of contextual data on forecasting performance. Although historical data lowers the variability of outcomes resulting from contextual data in processing linear cue-criterion relations, they heighten the variability of outcomes in interpreting nonlinear cue-criterion relations.

When relying on hierarchical regression[10] instead of MANOVA to re-analyze our data, the moderation effect is also reflected in the beta weights associated with the regression model. If we focus on U^2 as the dependent variable and convert beta values into relative weights, historical data account for 77% of the variance explained by the model. This result shows that nonlinear information processing primarily represents a function of the amount of historical data involved. Similarly, when computing such relative weights for G^2, historical data account for 78% of the model weights, whereas contextual data are associated with a relative beta weight of 32%. These results stress the importance of accounting for the detrimental effect of ambiguous

[10] This analysis is made possible by the fact that our predictor variables are only associated with two categorical levels.

historical data when generating judgmental forecasts about fashion products. Our empirical findings therefore provide support for Hypothesis H_{3b}.

Absolute Forecast Deviations. To further assess the meaningfulness of our findings for music forecasting practice, we focused on the extent to which providing or removing the two decision support types would impact forecasting effectiveness. More specifically, we started out by calculating absolute deviations of predictions from the actual chart positions for both best-fitting linear models and judges (Figure 4).

"Figure 4 here"

One-sided ANOVA tests indicated that all mean differences across decision support types were significant at the $p < 0.01$ level. In general, absolute forecast deviations were slightly larger for judges, ranging between 12.78 and 23.55 chart positions compared to 10.96 and 23.66 chart positions for the best-fitting linear model. For both the best-fitting linear model and judges the joint presence of historical and contextual data (C^+ / H^+) led to the smallest absolute deviations, whereas removing the two decision support anchors resulted in the largest deviation. However, in the latter case, judgmental forecasts resulted in slightly narrower prediction intervals than model predictions. We then constructed a series of hypothetical scenarios, in which we used our regression estimates to adjust absolute forecast deviations to simulate the removal or addition of a specific decision support anchor. Adjustments were made by accounting for the proportion of residual variance that was not explained by judges in each of the four conditions. Figure 5 provides an overview of our analysis. In particular, it indicates that the forecast deviations can be reduced on average by 4.28 chart positions when one additional decision support type is offered and 9.72 positions when both contextual and historical data are provided.

"Figure 5 here"

6. Discussion

We conducted this study to obtain a better understanding of the conditions that drive judgmental effectiveness when generating forecasts in the domain of fashion products. The dynamic nature of such environments makes it difficult to use conventional forecasting tools. We specifically focused on historical and contextual data, because they frequently form integral components of decision support systems. Our findings will help forecasters design better decision support systems, by outlining the conditions under which such data are likely to improve or reduce their ability to detect and interpret critical information.

Specifically, we show that when forecasters are concerned with predictive accuracy and only managerial judgments are employed, providing both historical and contextual data is beneficial. Moreover, our analyses indicate that if judgmental forecasts are combined with other methods, decision support provided to forecasters should be restricted to contextual anchors. In particular, we find that the exploitation of nonlinearities is easiest for human judgment if contextual, but no historical data are present. Thus, if the role of managerial judgment is to detect these nonlinearities (and the linearities are taken care of by some statistical model with which managerial judgments are combined), then a restriction of the decision support data provided to decision makers makes sense.

The music data this research builds on reinforce the conclusions of prior research demonstrating that contextual data make it harder for forecasters to perform on a par with the best-fitting linear model of the ecology (Karelaia and Hogarth 2008). We extend these findings by suggesting that, at the same time, the presence of contextual data may actually facilitate forecasters' ability to process nonlinearities. In addition, historical data significantly affect forecasters' effectiveness in analyzing nonlinear cue-criterion relations. In the absence of historical data, forecasters explain far more of the linear model's residual variance than when such information is available. Thus, our findings provide empirical support for Dane and Pratt's (2007) theoretical proposition that efficiently exploiting nonlinearities becomes more important as tasks become less predictable. We further investigated the interactive relationship between historical demand and contextual anchors and showed supporting evidence for the proposition that historical data moderate the effect of contextual anchors when analyzing linearities and nonlinearities.

Our empirical data indicate that forecasters' ability to exploit nonlinearities only fully unfolds when they can rely on contextual data and when judgmental biases due to misleading historical data are kept at a minimum. Prior research suggests that in these types of task environments, decision makers frequently rely on heuristic shortcuts such as pattern matching to help them process task information in a more efficient way (Gigerenzer 1999; Hoch and Schkade 1996). Because judgment analysis studies frequently model forecasts *as if* judges utilize a linear model (Hogarth and Karelaia 2007), one interesting path for future research could be to adopt Klein's (2003) Recognition-Primed Decision Models (RPDM) to analyze forecasters' underlying cognitive processes. RPDM assume that decision makers are capable of identifying cue patterns in highly uncertain environments because of a variety of *action scripts* developed over years of training. Typically, RPDM have been investigated in naturalistic decision-making settings in which actors (such as firefighters and emergency service units) have little time to process incomplete, highly uncertain task information (Weick 1988). The findings of our research warrant further studies of RPDM, which could systematically decompose the properties of a forecasting task and then employ more qualitative research designs to investigate how expert forecasters process relevant information to form mental representations of the underlying forecasting environment.

From the perspective of general systems theory (Churchman 1971; Flood and Carson 1988), our empirical findings may also provide a better understanding of the way judges tend to engage in individual learning. Cooksey (2000) argues that managerial decision behavior can be understood as a direct consequence of negative and positive feedback loops. Negative system feedback refers to the gap between a system's current position and the goal that it tries to attain. It triggers managerial actions aimed at reaching a level of convergence between current position and targeted goal. In contrast, positive system feedback may lead to a fundamental revision of managers' existing system goals by encouraging them to try out new paths that could result in the attainment of even higher-level objectives (Cooksey 2000). The underlying logic of negative and positive system feedback is therefore consistent with Argyris' (1990) notion of single (emphasizing negative feedback) and double (emphasizing positive feedback) loop learning. When considering forecasters' ability to process linear cue-criterion relations, we speculate that our observed interaction primarily triggers learning from negative system feedback. More specifically, the presence of both historical demand and contextual anchors will increase the need to take corrective measures to approximate the performance of the best-fitting linear model. Conversely, positive system feedback seems more likely to improve judges' effectiveness in processing nonlinear cue-criterion relations as forecasters need to identify and employ new organizing principles for improving forecasting accuracy. Our empirical results illustrate that learning from positive system feedback is most likely to occur in the absence of historical demand data, when the consideration of contextual knowledge leads to the greatest variability in judges' performance. We believe that our study opens the path to a number of potentially interesting research projects which could specifically test the relationship between learning from system feedback and forecasting effectiveness. Lastly, our findings have clear implications for the effective design of quick response policies in music companies (Cachon and Swinney 2011; Fisher and Raman 1996; Iyer and Bergen 1997). Specifically, as pop music singles are commonly used as pre-seasonal sales indicators of an upcoming full album release, efficiently predicting chart positions of music singles is likely to further reduce lead time by making performance signals become available even before the official publication of the music singles charts.

Additional research on the impact of historical demand and contextual anchors is crucially needed. Indeed, being able to employ the mode of information processing most likely to be effective in a given situation has recently been argued to represent a key skill in managerial decision-making (Seifert and Hadida 2013). Our findings shed further light on the circumstances that favor reliance on model prediction or expert judgment, particularly since many organizations utilize linear model forecasts as a fundamental tool for coping with uncertainty. Prior research in this domain suggested that forecast combinations relying on a simple 50:50 split between model and manager judgment efficiently exploited nonlinear cue-criterion relations in the task environment (Blattberg and Hoch 1990). By specifically demonstrating how the presence of different types of decision support anchors can harm forecasters' ability to interpret informational cues, our research provides a more refined view of the drivers behind judgmental performance.

To conclude, we hope that future research will provide additional tests of the robustness of our results by adopting alternative research designs, prediction tasks, and operationalizations of key constructs. Specifically, the extent to which the reported relationship between task-level characteristics and judgmental performance is robust and extends to other types of forecasting contexts involving fashion products (e.g., the textile or movie industries) remains unclear. Despite our efforts to reduce systematic error in our chosen methodology, our operationalization of forecasters' nonlinear information processing skill could not fully rule out the possibility that omitted variables contaminated our findings. We therefore encourage replication of our study in alternative contexts. Moreover, further insight into the effectiveness of judges' sense-making processes may also be gained by developing more sophisticated models that incorporate key nonlinearities from the task environment (such as the visual appearance or sound sample in our study). Finally, future research could examine the extent to which the "wisdom of the crowd" associated with collective forecasting judgments may compensate for the performance limitations of group members when facing different types of task characteristics. In sum, we believe that the results of our study raise a number of intriguing questions for follow-up research, and offer promising opportunities to further clarify the role of decision support systems in judgmental forecasting.

References

Andreassen, P. B., and Kraus, S. J. (1990). Judgmental extrapolation and the salience of change. *Journal of Forecasting*, 9(4), 347-372.

Argyris, C. (1990). *Overcoming Organizational Defenses: Facilitating Organizational Learning*, Needham Heights, MA: Allyn and Bacon.

Au, K. F., Choi, T. M., and Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. *International Journal of Production Economics*, 114(2), 615-630.

Bendoly, E., Donohue, K., & Schultz, K. L. (2006). Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management,* 24(6), 737-752.

Blattberg, R. C., Hoch S. J. (1990). Database models and managerial intuition: 50% model + 50% manager. *Management Science* 36, 887-899.

Bolton, G. E., & Katok, E. (2008). Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management*, 10(3), 519-538.

Bostian, A. A., Holt, C. A., & Smith, A. M. (2008). Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management*, 10(4), 590-608.

Boulaksil, Y., P. H. Franses. (2009). Experts' stated behavior. *Interfaces*, 39(2): 168-171.

Brehmer, B. (1978). Response consistency in probabilistic inference tasks. *Organizational Behavior and Human Performance*, 22(1), 103-115.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: The University of California Press.

Cachon, G. P., and Fisher, M. (2000). Supply chain inventory management and the value of shared information. *Management Science*, 46(8), 1032-1048.

Cachon, G. P., & Swinney, R. (2011). The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. *Management Science*, 57(4), 778-795.

Camerer, C., Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5, 325-370.

Choi, T. M., Hui, C. L., Liu, N., Ng, S. F., and Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, 59, 84-92.

Choi, T. M. J., Li, D., & Yan, H. (2006). Quick response policy with Bayesian information updates. *European Journal of Operational Research,* 170(3), 788-808.

Christopher, M., Lowson, R., and Peck, H. (2004). Creating agile supply chains in the fashion industry. *International Journal of Retail and Distribution Management*,32(8), 367-376.

Churchman, C.W. (1971). *The Design of Inquiring Systems*. Basic Books Inc., Perseus Books Group.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5, 559-583.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic.

Cohen. J. Cohen, I. (1975). *Applied Multiple Regression/Correlation Analysis of the Behavioral Sciences*, Erlbaum, Hillsdale. NJ.

Cooksey, R. (1996). *Judgment Analysis: Theory, Methods and Applications*. NY: Acad Press.

Cooksey, R.W. (2000). Mapping the texture of managerial decision making: A complex dynamic decision perspective. *Emergence: A Journal of Complexity Issues in Organizations and Management*, 2,102-122.

Dane, E., Pratt M. G. (2007). Exploring intuition and its role in managerial decision making. *Academy of Management Review,* 32, 33-54.

Dhami, M. K. and Harries, C. (2001). Fast and frugal versus regression models of human judgment. *Thinking and Reasoning*, 7, 5-27.

Dougherty, M. R. and Thomas, R. P. (2012). Robust decision making in a nonlinear world. *Psychological Review*, 119(2), 321-344.

Dunwoody, P. T., Haarbauer, R., Marino, C., Tang, C. (2000). Cognitive adaptation and its consequences: A test of cognitive continuum theory. *Journal of Behavioral Decision Making,* 13, 35-54.

Einhorn, H. J. (1970). The use of nonlinear noncompensatory models in decision making. *Psychological Bulletin*, 73, 221–230.

Einhorn, H. J. (1971). Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance*, 6, 1-27.

Einhorn, H. J. (1974). Cue Definition and Residual Judgment. *Organizational Behavior and Human Decision Processes*, 12, 30-49.

Einhorn, H. J. and Hogarth, R. M. (1975). Unit weighting schemes for decision-making. *Organizational Behavior and Human Decision Processes*, 13, 171-192.

Elberse, A., and Eliashberg, J. (2003). Demand and supply dynamics of sequentially released products in international markets: The case of motion pictures. *Marketing Science*, 22(3), 329-354.

Eliashberg, J., and Sawhney, M. S. (1994). Modeling goes to Hollywood: Predicting individual differences in movie enjoyment. *Management Science*,40(9), 1151-1173.

Eliashberg, J., Weinberg, C. B., and Hui, S. K. (2008). Decision models for the movie industry. In *Handbook of marketing decision models* (pp. 437-468). Springer US.

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191.

Fellner, W. (1961), Distortion of subjective probabilities as a reaction to uncertainty. *Quarterly Journal of Economics*, 75, 670-694.

Fisher, M., & Raman, A. (1996). Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research*, 44(1), 87-99.

Flood, R.L., Carson, E.R. (1988) *Dealing with Complexity: An Introduction to the Theory and Application of Systems Science,* New York: Plenum.

Frisch, D., Baron, J. (1988). Ambiguity and rationality, *Journal of Behavioral Decision Making*, 1, 149-157.

Gaur, V., Kesavan, S., Raman, A., and Fisher, M. L. (2007). Estimating demand uncertainty using judgmental forecasts. *Manufacturing and Service Operations Management*, 9(4), 480-491.

Gigerenzer, G., Todd, P.M., ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford University Press, Oxford, UK.

Goodwin, P. and Wright, G., (1993). Improving judgmental time series forecasting: a review of the guidance provided by research, *International Journal of Forecasting*, 9(2) 147-161.

Green, M., and Harrison, P. J. (1973). Fashion forecasting for a mail order company using a Bayesian approach. *Journal of the Operational Research Society*, 24(2), 193-205.

Gutierrez, R. S., Solis, A. O., and Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111(2), 409-420.

Hadida, A.L. and Paris, T. (2014). Managerial Cognition and the Value Chain in the Digital Music Industry. *Technological Forecasting and Social Change*, 83: 84-97.

Hamm, R. M. (1987). Clinical intuition and clinical analysis: expertise and the cognitive continuum. In Jack Dowie and Arthur Einstein (Eds), *Professional Judgment*, Cambridge University Press.

Hammond K. R. (1996). *Human judgment and social policy*. New York: OUP.

Hammond, K. R., and Summers, D. A. (1972). Cognitive control. *Psychological Review, 79*, 58–67.

Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, 63(3), 247-263.

Harvey, N., Ewart, T., and West, R. (1997). Effects of data noise on statistical judgement. *Thinking & Reasoning*, 3(2), 111-132.

Hines, T., and Bruce, M. (Eds.). (2007). *Fashion marketing*. Routledge.

Hirsch, P. M. (1972). Processing fads and fashions: An organization-set analysis of cultural industry systems. *American Journal of Sociology*, 77(4), 639-659.

Ho, T.-H., Lim, N., Cui, T.H., (2010). Reference dependence in multilocation newsvendor models: A structural analysis. *Management Science*, 56, 1891–1910.

Hoch, S. J., and Schkade, D. A. (1996). A psychological approach to decision support systems. *Management Science*, 42(1), 51-64.

Hogarth, R. M. (1987). *Judgment and choice*. 2nd edition. New York: Wiley.

Hogarth, R. M., and Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, 114(3), 733-58.

IFPI; 2014. http://www.ifpi.org.

Iman, R. L., and Conover, W. J. (1979). The use of the rank transform in regression. *Technometrics*, 21(4), 499-509.

Iyer, A. V., & Bergen, M. E. (1997). Quick response in manufacturer-retailer channels. *Management Science*, 43(4), 559-570.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237.

Karelaia, N., and Hogarth, R.M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, 134, 404–426.

Klein, G. (2003). *Intuition at work*. New York: Doubleday.

Kremer, M., Moritz, B., and Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, 57(10), 1827-1843.

Lattin, J. M., Carroll, J. D., and Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA, USA: Thomson Brooks/Cole.

Lawrence, M., Goodwin, P., O'Connor, M., and Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22, 493-518.

Lawrence, M., and Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, 43(2), 172-187.

Lee, Y.S. and Siemsen, E. (2015). Task Decomposition and Newsvendor Decision Making. *Working Paper*. Available at SSRN: http://ssrn.com/abstract=2232073.

Lee, J. W., Yates, J. F. (1992). How quantity judgment changes as the number of cues increases: An analytical framework and review. *Psychological Bulletin*, 112, 363-377.

Mentzer, J.T. and M.A. Moon (2005). *Sales Forecasting Management*. 2nd Edition, Sage Publications: London, UK.

Moe, W. W., and Fader, P. S. (2001). Modeling hedonic portfolio products: A joint segmentation analysis of music compact disc sales. *Journal of Marketing Research*, 376-385.

Moreau, F., and Peltier, S. (2004). Cultural diversity in the movie industry: A cross-national study. *Journal of Media Economics*, 17(2), 123-143.

Nenni, M. E., Giustiniano, L., and Pirolo, L. (2013). Demand forecasting in the fashion industry: a review. *International Journal of Engineering Business Management*, 5(37), 1-6.

Ni, Y., and Fan, F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. *Expert Systems with Applications*, 38(3), 1529-1536.

Olshavsky, R. (1979) Task complexity and contingent processing in decision-making: A replication and extension. *Organizational Behavior and Human Performance*, 24, 300- 316.

Payne, J. (1976) Task complexity and contingent processing in decision-making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.

Peterson, R. A., and Berger, D. G. (1971). Entrepreneurship in Organizations: Evidence from the Popular Music Industry. *Administrative Science Quarterly*, 16(1), 97-106.

Ramsey, J.B. (1969) Tests for Specification Errors in Classical Linear Least Squares Regression Analysis, *Journal of the Royal Statistical Society*, Series B., 31(2), 350–371.

Ren, Y., Croson, R., (2013). Overconfidence in Newsvendor Orders: An Experimental Study. *Management Science*, 59 2502–2517.

Sanders, N. R., K. B. Manrodt. (2003). Forecasting software in practice: Use, satisfaction, and performance. *Interfaces*, 33(5): 90-93.

Sani, B., and Kingsman, B. G. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48(7), 700-713.

Santagata, W. (2004). Creativity, fashion and market behavior. In Power, D., and Scott, A. J. *Cultural industries and the production of culture*. Routledge. 75-90.

Sawhney, M. S., and Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, 15(2), 113-131.

Schweitzer, M. E., & Cachon, G. P. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science,* 46(3), 404-420.

Serel, D. A. (2009). Optimal ordering and pricing in a quick response system. *International Journal of Production Economics*, 121(2), 700-714.

Sichel, B. (2008). Forecasting Demand with Point of Sales Data: A Case Study of Fashion Products. *Journal of Business Forecasting Methods and Systems*, 27(4), 15-16.

Seifert, M., Hadida, A.L. (2013). On the Relative Importance of Linear Model and Human Judge(s) in Combined Forecasting. *Organizational Behavior and Human Decision Processes*, 120, 24-36.

Slovic, P., and Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6, 649-744.

Snyder, R. (2002). Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research*, 140(3), 684-699.

Steenkamp, J.-B. E. M., F. Hofstede, M. Wedel (1999), A cross-national investigation into the individual and national cultural antecedents of consumer innovativeness. *Journal of Marketing*, 63, 55-69.

Steinmann, D. (1976). The effects of cognitive feedback and task complexity in multiple-cue probability learning. *Organizational Behavior and Human Performance*, 15, 168-179.

Stewart, T. R. (2001). The lens model equation. In K. R. Hammond and T. R. Stewart (Eds.), *The Essential Brunswik: Beginnings, Explications, Applications* (pp. 357-362). New York: Oxford University Press.

Stewart, T. R., Roebber, P. J., and Bosart, L. F. (1997). The Importance of the Task in Analyzing Expert Judgment. *Organizational Behavior and Human Decision Processes*, 69(3), 205-219.

Stremersch, S., G. J. Tellis. (2004). Understanding and managing international growth of new products. *International Journal of Research in Marketing,* 21, 421-438.

Su, X. (2008). Bounded Rationality in Newsvendor Models. *Manufacturing & Service Operations Management*, 10, 566–589.

Sun, Z. L., Choi, T. M., Au, K. F., and Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1), 411-419.

Taylor, J. W., and Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 19(1), 57-70.

Thomassey, S., Happiette, M., and Castelain, J. M. (2005). A global forecasting support system adapted to textile distribution. *International Journal of Production Economics*, 96(1), 81-95.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. *Psychological Review,* 71(6), 528-530.

Van den Bulte, C., S. Stremersch. (2004). Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science,* 23(4), 530-544.

Vogel, H. L. (2007). *Entertainment Industry Economics: A Guide for Financial Analysis (7th Edition).* Cambridge: Cambridge University Press.

Weick, K. (1988). Enacted Sensemaking in Crisis Situation, *Journal of Management Studies*, 25(4), 305–317.

Wong, W. K., and Guo, Z. X. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*,128(2), 614-624.

Wood, R.E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes,* 37, 60-82.

Xia, M., Zhang, Y., Weng, L., and Ye, X. (2012). Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. *Knowledge-Based Systems*, 36, 253-259.

Yu, Y., Choi, T. M., and Hui, C. L. (2011). An intelligent fast sales forecasting model for fashion products. *Expert Systems with Applications*, 38(6), 7373-7379.

Zuckerman, E.W., and Kim, T.Y. (2003). The critical trade-off: Identity assignment and box office success in the feature film industry. *Industrial and Corporate Change*, 12(1), 27-67.

Tables and Figures

Table 1 Overview of stimuli included in the study

		Decision Support Type			
Predictor Type	*Level of Analysis*	*C-/H-*	*C-/H+*	*C+/H-*	*C+/H+*
Artist					
Type	**Band versus single artist?**	✓	✓	✓	✓
Online fanbase	**Log of *Myspace* members**	✓			
Prior recognition	**No of sales awards received**		✓		✓
Chart history: album	**Mean chart position**		✓		✓
Chart history: single	**Mean chart position**		✓		✓
Product					
Music genre	**Mainstream pop/rock? (Y/N)**	✓	✓	✓	✓
Album	**Has an album been released? (Y/N)**	✓	✓	✓	✓
Producer	**No of previous Top 100 hits**	✓	✓	✓	✓
International success	**Included US charts? (Y/N)**	✓	✓	✓	✓
Radio	**Highest airplay chart position**			✓	✓
Printed media	**Mean rating in key press reviews**			✓	✓
Television	**Highest music video chart position**			✓	✓
Marketing	**Log of expenditure in US dollars**	✓		✓	✓

"✓" indicates that predictor variable was included in best fitting linear model

C^-: Contextual anchor absent C^+: Contextual anchor present

H^-: Historical data anchor absent H^+: Historical data anchor present

Table 2 Characteristics of decision support conditions

	Decision Support Type			
Measurement	C^- / H^-	C^- / H^+	C^+ / H^-	C^+ / H^+
Number of predictors included in model	7	8	9	12
Intercorrelation of predictors (r)	-0.03	0.04	-0.03	0.11
Standard deviation of β weights	3.71	3.25	3.88	2.45
Sample size (N)	43	47	59	61

C^-: Contextual anchor absent C^+: Contextual anchor present

H^-: Historical data anchor absent H^+: Historical data anchor present

Intercorrelations and standard deviations are averaged across all pairs of predictor variables

Table 3 Summary of regression results

Decision Support Type	Best Linear Fit Model* Model accuracy ($R_{O,X}$)	Linear model Fit ($\hat{M}$)	Unexplained variance	Judgment Model** Judgment accuracy ($R_{Y,O}$)	Linear model Fit ($\hat{M}^*$)	Unexplained variance	Judgment Decomposition Exploiting linearities (G^2)	Exploiting nonlinearities (U^2)
C^- / H^-	.53	.28	.72	.57 (.09)***	.21	.68	.92	.11
C^- / H^+	.71	.50	.50	.72 (.11)	.46	.48	.96	.06
C^+ / H^-	.62	.38	.62	.68 (.12)	.31	.53	.93	.16
C^+ / H^+	.83	.68	.32	.78 (.06)	.52	.40	.84	.08

Contextual data anchor: C^+ = present, C^- = absent; Historical data anchor: H^+ =present, H^- = absent
*Cross-validated results
**Mean values across participants
***Standard deviation in parentheses

Figure 1 Overview of the Judgment Analysis approach (adapted from Cooksey 1996)

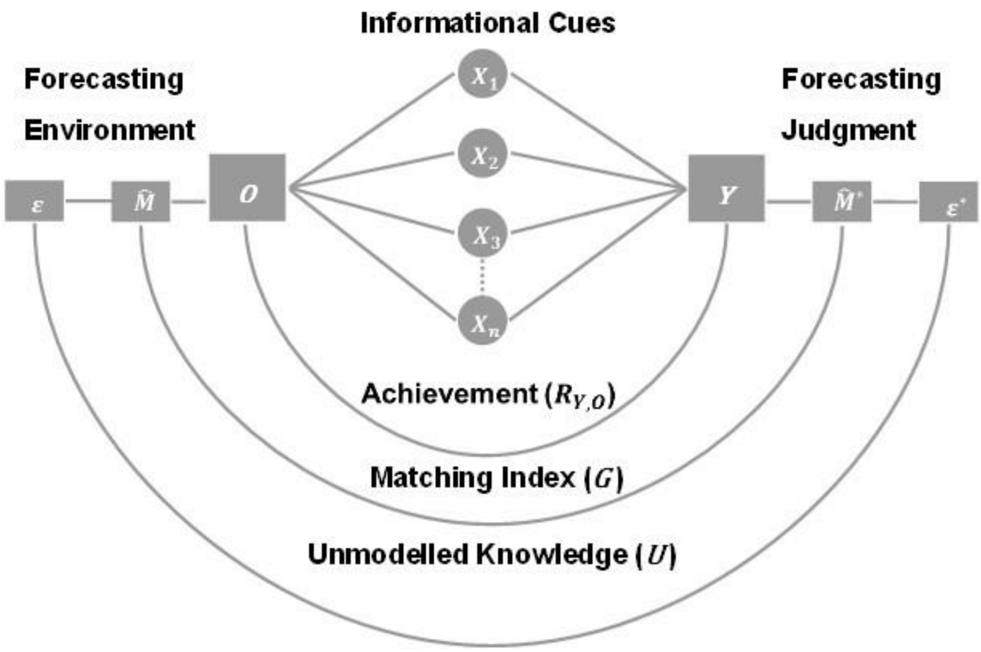


Figure 2 Correlations of linear and nonlinear forecasting efficiency (*G*U*)

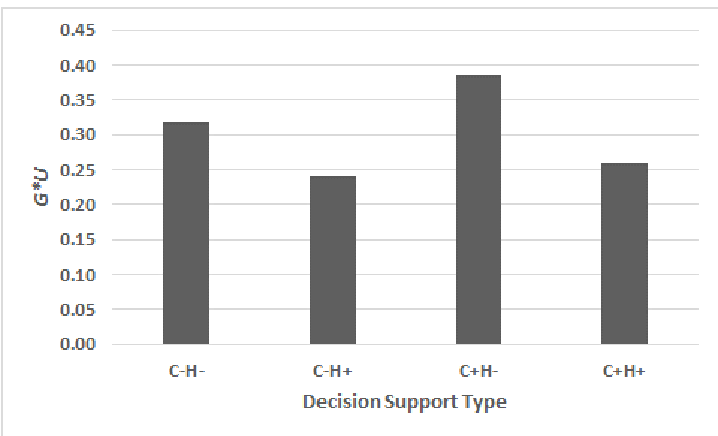


Figure 3 Effect of contextual anchors and historical demand anchors on judges' effectiveness in exploiting linear and nonlinear cue-criterion relations

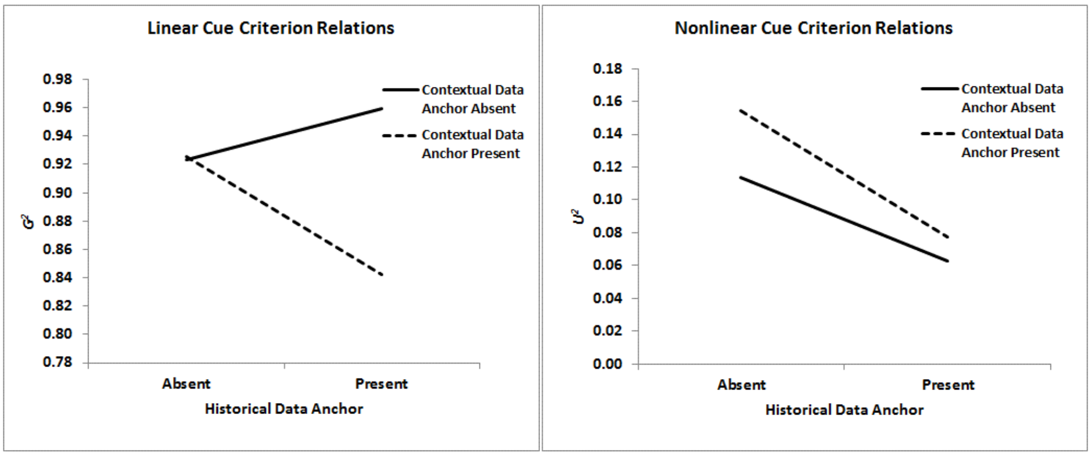
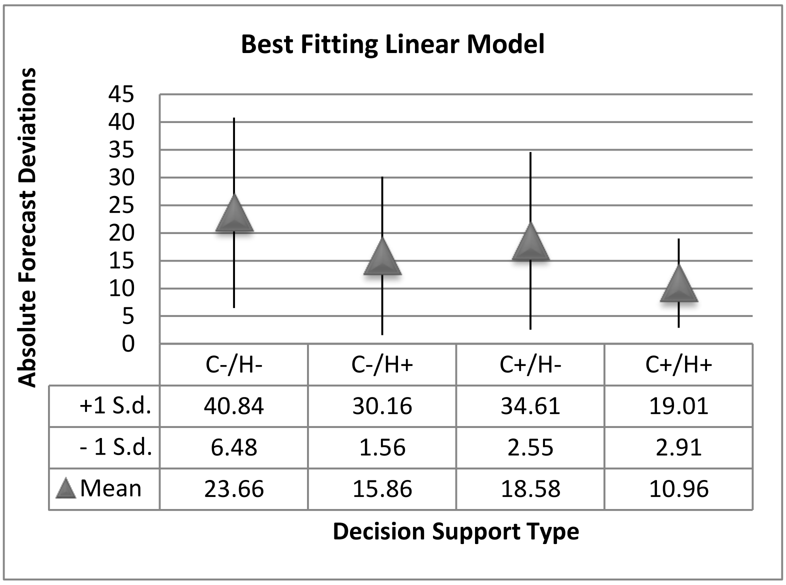
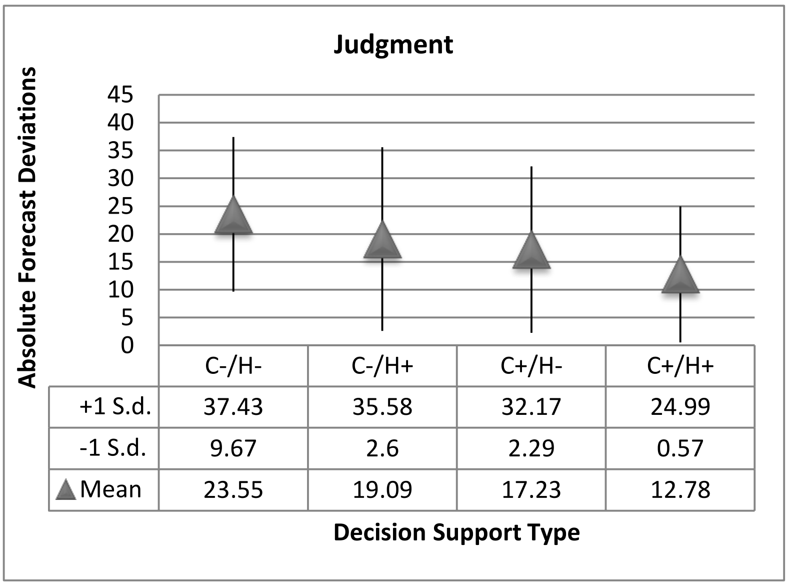


Figure 4 Absolute deviations from forecasting event

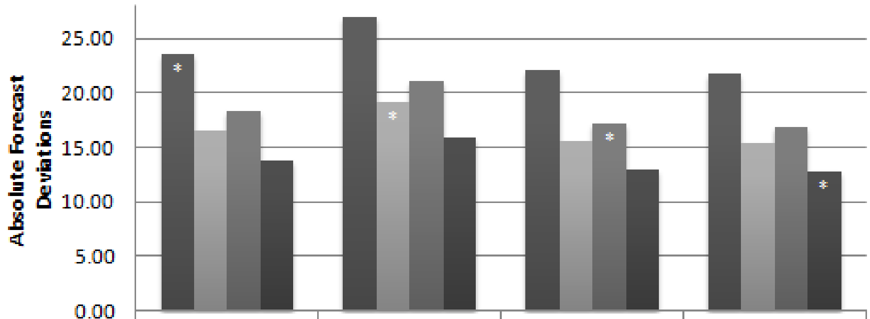


	C-/H-	C-/H+	C+/H-	C+/H+
+1 S.d.	40.84	30.16	34.61	19.01
- 1 S.d.	6.48	1.56	2.55	2.91
▲Mean	23.66	15.86	18.58	10.96



	C-/H-	C-/H+	C+/H-	C+/H+
+1 S.d.	37.43	35.58	32.17	24.99
-1 S.d.	9.67	2.6	2.29	0.57
▲Mean	23.55	19.09	17.23	12.78

Figure 5 Forecasting effectiveness with and without contextual and historical data anchors ("*" indicates actual deviation)



	C-/H-	C-/H+	C+/H-	C+/H+
C-/H-	23.55*	27.04	22.11	21.73
C-/H+	16.62	19.09*	15.60	15.34
C+/H-	18.36	21.08	17.23*	16.93
C+/H+	13.85	15.91	13.00	12.78*

Appendix

Two examples from our data collection instrument are provided below. While for the first single neither contextual nor historical demand anchors are available (i.e. a new artist who has not been promoted intensively prior to the single's release), in the second example a contextual anchor is present as an already established artist has been promoted through different media channels.

ARTIST BACKGROUND:

Started in:	2006
Country:	Germany
Previous awards:	-
Previous singles:	-
Previous albums:	-
US/UK chart success (average):	Album: -/ - Singles: - / -
Myspace blog members:	2625

TITLE:	GOODBYE (sound sample)	Top 100 Airplay charts:	- (full table here)
ARTIST:	MCKENNA	US/UK Single charts:	-/-
Label:	Deutsche Ton	MTV/VIVA video charts:	-/-
Release:	09/03	Mean rating in music press:	-
Producer/Songwriter:	McKenna/McKenna	Retail promotion campaign launched?	No
Pre-order price:	5.99 Euros	Marketing expenses:	8.500 Euros

Single 6 out of 25

ARTIST BACKGROUND:

Started in:	1993
Country:	Canada
Previous awards:	3x Platinum
Previous singles:	7
Previous albums:	3
US/UK chart success (average):	Album: #7/ #18 Singles: #12/ -
Myspace blog members:	31.625

TITLE:	FALLEN LEAVES (sound sample)	Top 100 Airplay charts:	#6 (full table here)
ARTIST:	BILLY TALENT	US/UK Single charts:	-/-
Label:	Atlantic Records	MTV/VIVA video charts:	11/17
Release:	09/03	Mean rating in music press:	2 of 5 stars
Producer/Songwriter:	Brown/Talent	Retail promotion campaign launched?	Yes
Pre-order price:	6.99 Euros	Marketing expenses:	32.3110 Euros