The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/1757-5818.htm

Received 10 December 2013
Revised 7 February 2014
26 April 2014
22 June 2014
Accepted 6 July 2014

Perceptions are relative

An examination of the relationship between relative satisfaction metrics and share of wallet

Timothy Lee Keiningham
Ipsos Loyalty, Parsippany, New Jersey, USA
Bruce Cooil
Graduate School of Management, Vanderbilt University, Nashville, Tennessee, USA
Edward C. Malthouse
Department of Integrated Marketing Communication, Northwestern University, Evanston, Illinois, USA
Alexander Buoye and Lerzan Aksoy
Schools of Business, Fordham University, Bronx, New York, USA, and
Arne De Keyser and Bart Larivière
Department of Management, Innovation and Entrepreneurship, Ghent University, Ghent, Belgium

Abstract

Purpose – There is general agreement among researchers and practitioners that satisfaction is relative to competitive alternatives. Nonetheless, researchers and managers have not treated satisfaction as a relative construct. The result has been weak relationships between satisfaction and share of wallet in the literature, and challenges by managers as to whether satisfaction is a useful predictor of customer behavior and business outcomes. The purpose of this paper is to explore the best approach for linking satisfaction to share of wallet.

Design/methodology/approach – Using data from 79,543 consumers who provided 258,743 observations regarding the brands that they use (over 650 brands) covering 20 industries from 15 countries, various models such as the Wallet Allocation Rule (WAR), Zipf-AE, and Zipf-PM, truncated geometric model, generalization of the WAR and hierarchical regression models are compared to each other.

Findings – The results indicate that the relationship between satisfaction and share of wallet is primarily driven by the relative fulfillment customers perceive from the various brands that they use (as gauged by their relative ranked satisfaction level), and not the absolute level of satisfaction.

Practical implications – The findings provide practical insight into several easy-to-use approaches that researchers and managers can apply to improve the strength of the relationship between satisfaction and share of wallet.

Originality/value – This research provides support to the small number of studies that point to the superiority of using relative metrics, and encourages the adoption of relative satisfaction metrics by the academic community.

Keywords Customer behaviour, Consumer satisfaction

Paper type Research paper

Journal of Service Management
Vol. 26 No. 1, 2015
pp. 2-43

1757-5818
DOI 10.1108/JOSM-12-2013-0345

Arne De Keyser and Bart Larivière acknowledge support from the Special Research Fund (BOF, Ghent University, project 01N04011) and the National Bank of Belgium. Bruce Cooil acknowledges support from the Dean's Fund for Faculty Research, Owen Graduate School, Vanderbilt University.

Managers widely believe that customer satisfaction is a fundamental determinant of long-term consumer behavior (Oliver, 1980; Yi, 1990). This widespread acceptance has made customer satisfaction the most widely used metric in the measurement and management of consumer loyalty (Aksoy, 2013a; Zeithaml *et al.*, 2006). Companies spend substantial amounts of money to measure and manage customer satisfaction. For example, Inside Research (2012) found that for the 13 marketing research firms that responded to their survey, revenue from US-only customer satisfaction research would exceed $750 million – this figure likely underestimates the total spent with marketing research firms given the small number of responding firms.

A review of the scientific literature on customer satisfaction supports management's focus on customer satisfaction. In particular, many studies have linked customer satisfaction to customers' purchasing behaviors (e.g. Bolton, 1998; Mittal and Kamakura, 2001; Rust and Zahorik, 1993). A close examination of the research regarding customer satisfaction and customers' share of category spending, however, reveals that while the relationship is positive, it tends to be very weak (Hofmeyr *et al.*, 2008; Mägi, 2003).

Because of the weak relationship, managers have difficulty connecting their efforts to improve satisfaction with tangible financial outcomes (Aksoy, 2013a; Keiningham *et al.*, 2014). For example, in an article entitled "Proof that it pays to be America's most-hated companies," *Bloomberg Businessweek* reported that their analysis of the relationship between the American Customer Satisfaction Index (ACSI) and stock performance found the relationship to be negative. Specifically, the magazine reported (Chemi, 2013):

> […] customer-service scores have no relevance to stock market returns […] the most-hated companies perform better than their beloved peers […] Your contempt really, truly doesn't matter […] If anything, it might hurt company profits to spend money making customers happy.

Results such as these have led to calls by some managers and researchers to discontinue the measurement and management of satisfaction (Gupta and Zeithaml, 2006). Books like *Customer Satisfaction is Worthless, Customer Loyalty is Priceless,* by consultant Jeffery Gitomer (1998), and articles like "Customer satisfaction: it is dead, but it will not lie down," by researchers Williams and Visser (2002) are indicative of this general frustration.

Given customer satisfaction's weak relationship to business outcomes and customer behaviors, Mägi (2003, p. 104) argues "it might be informative to use relative measures of satisfaction when predicting customer share" (i.e. share of wallet). Researchers agree that perceptual metrics such as satisfaction need to be measured relative to competitive alternatives (e.g. Varki and Rust, 1997). Furthermore, there is a large body of research confirming the influence of competitive comparisons on both choice and post-purchase evaluations (e.g. Rust *et al.*, 2000).

The small number of studies that have used relative satisfaction in the scientific literature (e.g. Bolton *et al.*, 2000; Bowman and Narayandas, 2004; Hardie *et al.*, 1993; Wind, 1970) point to the superiority of relative metrics in linking to customer behavior. Nonetheless, the scientific community has been slow to use relative satisfaction in their research. None of the methods used by these researchers have been widely used in other scientific investigations. Rather, the overwhelming majority of scientific research investigating satisfaction relies on absolute metrics on a single firm. Furthermore, these methods are rarely used by managers.

The same reluctance to use relative metrics cannot be said for the practitioner community. Some of the world's largest survey research organizations specifically advocate the use of relative metrics when linking customer satisfaction to a customer's share of wallet, and make them the foundation of their brand equity and customer experience measurement approaches, i.e., TNS: Conversion Model (Louw and Hofmeyr, 2012), Ipsos: Brand Value Creator (Hofmeyr *et al.*, 2008), and Ipsos: Wallet Allocation Optimizer (Keiningham *et al.*, 2011). These firms report strong correlations between their approaches and share of wallet.

The creators of these frameworks have made them widely available for managers to apply in their organizations by publishing their methodologies. Each of these approaches, however, uses a different technique to link relative metrics to share of wallet. Furthermore, despite their publication, these methodologies are not often used by managers outside of their application within a research firm's specific product offer. This, however, does not mean that they are not widely used. For example, the Conversion Model is used by "80% of the world's most valuable brands" (TNS, 2012).

The gap between the science and the practice of marketing in this regard has profound implications for both managers and researchers. There is no research in the peer reviewed literature that rigorously investigates various methodologies to determine their efficacy. As a result, researchers and managers are left with almost no guidance as to the usefulness of different approaches, or even to the validity of relative satisfaction metrics in general.

Additionally, if relative metrics more accurately reflect the relationship between satisfaction and customers' share of category spending, this would likely serve as impetus for new research in a number of areas. Clearly, this would necessitate new research into the relative nature of satisfaction and its corresponding impact on consumer behavior. It would also likely spur examinations into the potential relative impact of other perceptual and attitudinal metrics on consumer behavior (e.g. commitment, emotions, etc.).

As a result, there is a need for research regarding the efficacy of relative satisfaction metrics and best practices regarding the use relative satisfaction metrics. This research fills these gaps by investigating the relationship between relative satisfaction and customers' share of category spending (i.e. share of wallet) using data from 79,543 consumers who provided 258,743 observations regarding the brands that they use within a particular industry category. Data included ratings of over 650 brands in 20 industries from 15 countries.

The results of this investigation find that relative satisfaction significantly outperforms absolute satisfaction levels in linking to customers' share of category spending. Models based upon absolute satisfaction levels were consistently the worst performing models investigated. Moreover, we find that the most commonly used power laws in practice perform well compared to other models investigated in linking relative satisfaction to share of wallet. Finally, we note that there are significant differences in the complexity of the various approaches examined. Therefore managers need to consider the trade-off between relationship strength and complexity when selecting the best approach for use within their firms.

Structure of manuscript

This investigation relies upon a rigorous investigation of different power laws and hierarchical regression models. As a result, a thorough description of the investigation requires a detailed presentation of several models and analytic procedures. This has the potential to make the paper quite technical and fragmented, resulting in a paper that is difficult for most managers to read. As a result, we believe that the core message of the paper can be lost in the technical descriptions of the models and analytics.

Therefore, in an effort to maximize the readability and insights gleaned from this investigation, this paper is divided into two main sections. The first section focuses on the theoretical foundation, core findings, and implications of the research. The second section is a Technical Appendix that provides a detailed overview of the models examined, and the various approaches used to investigate the properties of these models.

By using this approach, we hope that we are able to provide researchers and managers with clear and relevant insights while maintaining scientific rigor and transparency regarding our analyses and findings.

Theoretical background

Customer satisfaction

Satisfaction is the consumer's emotional response to the fulfillment of needs, expectations, wishes or desires. Specifically, Oliver (2010, p. 8) defines customer satisfaction as follows: "Satisfaction is the consumer's fulfillment response. It is a judgment that a product/service feature, or the product or service itself, provided (or is providing) a *pleasurable* level of consumption-related fulfillment, including levels of under- or overfulfillment."

Researchers have extensively examined the theoretical underpinnings of the satisfaction construct (e.g. Fornell *et al.*, 1996; Luo and Bhattacharya, 2006; Oliver, 1997). Researchers have also investigated the effects of customer satisfaction on future consumer behaviors (e.g. Crosby and Stephens, 1987; Keiningham *et al.*, 2003; Luo and Homburg, 2007).

Of particular importance to this investigation, there is general agreement among researchers and practitioners that satisfaction is relative to perceived competitive alternatives (e.g. Birtchnell, 1994; Holt and Huber, 1969; Varki and Rust, 1997; Semon, 1994). For example, Woodruff *et al.* (1983) argue that norms based on consumer experiences with brands within a product category and relative to competing alternatives in that category were a more natural comparison standard than focal brand expectations. Research by Cadotte *et al.* (1987) found that experience-based norms better explain variations in satisfaction than focal brand expectations. Additionally, Gardial *et al.* (1994) found that consumers tend to rely on competitive comparisons/norms when evaluating their consumption experiences.

This can in part be explained by expectancy-disconfirmation model of the appraisal sequence for satisfaction (Oliver, 2010, pp. 355-360). Oliver (2010, p. 22) defines expectancy-disconfirmation as "the psychological interpretation of an expectation-performance discrepancy. Consumers would describe this concept in terms of the performance of a product or service being better or worse than expected."

Although satisfaction and disconfirmation are not perfectly correlated, "satisfaction results primarily from disconfirmation" (Rust *et al.*, 1996, p. 234). As such, expectations tend to play a strong role in consumers' satisfaction judgments.

Consumers' expectations are strongly affected by their experiences. Experiences, however, are not limited to the focal/purchased brand, but frequently include broader experiences within a product or service category (Woodruff *et al.*, 1983). In addition, expectations may be affected by advertising and word of mouth (Boulding *et al.*, 1993; Miller, 1977). This, to a large degree, explains why satisfaction is influenced by competitive comparisons or norms.

Customer satisfaction and share of wallet

The relationship between satisfaction and consumer behavior is grounded in the theory of planned behavior (Ajzen and Madden, 1986), an offshoot of the theory of reasoned action (Ajzen, 2001; Ajzen and Fishbein, 1980). The theory argues that behaviors are influenced by three factors: attitudes, subjective norms, and perceived behavioral control. Specifically, favorable/unfavorable attitudes, in combination perceived societal "norms" are the primary determinants of a consumer's intention to perform a behavior (provided the consumer believes he/she has the ability to perform the behavior). Although satisfaction is generally viewed as a perception (e.g. Oliver, 1980) this reflects the generally accepted view of how satisfaction ultimately influences consumer purchase decisions (Mittal and Frennea, 2010).

Share of wallet is widely believed to be driven in part by customers' perceptions of the brands they use. The chain of effects can be thought of as product/service performance → satisfaction → share of wallet. In fact, this chain of effects is a logical adaptation of the core chain of effects proposed in some of the seminal models in marketing (Keiningham *et al.*, 2005): SERVQUAL (Parasuraman *et al.*, 1988; Zeithaml *et al.*, 1996), service profit chain (Heskett *et al.*, 1994), return on quality (Rust *et al.*, 1995), and the satisfaction profit chain (Anderson and Mittal, 2000).

The idea that customer satisfaction should link to share of category spending is intuitive (i.e. we tend to spend more with firms that better satisfy us). A large body of research does support this positive relationship (e.g. Baumann *et al.*, 2005; Bowman and Narayandas, 2004; Cooil *et al.*, 2007; Keiningham *et al.*, 2003, 2005; Larivière, 2008; Mägi, 2003; Perkins-Munn *et al.*, 2005; Silvestro and Cross, 2000).

The problem from a managerial perspective, however, is that while there tends to be a statistically significant positive relationship between satisfaction and share of wallet, the percentage of variance explained by this relationship is low (Hofmeyr *et al.*, 2008; Mägi, 2003). As a result, managers have openly challenged "whether the relationship between unobservable measures such as customer satisfaction and observable behavior such as purchasing was sufficiently strong to justify its use as the primary unobservable predictor" (Gupta and Zeithaml, 2006, p. 721).

Researchers have proposed two possible reasons to explain this weak relationship. First, customers appear to differ in their sensitivity to variations in satisfaction (Hofmeyr and Parton, 2010). For example, demographic differences have been shown to impact the satisfaction-share of wallet relationship (Cooil *et al.*, 2007). Second, researchers argue that satisfaction's impact on customer behavior is nonlinear and asymmetric (e.g. Anderson and Mittal, 2000; Crotts *et al.*, 2008; Keiningham and Vavra, 2001). Accounting for the asymmetric, non-linear pattern of satisfaction has improved the relationship between satisfaction and share of wallet (e.g. Bowman and Narayandas, 2004; Keiningham *et al.*, 2003). Nonetheless, a large portion of the variance remains unexplained (Hofmeyr and Parton, 2010).

An alternative explanation for the weak relationship has been proposed by members of the practitioner community. Hofmeyr and Parton (2010) argue that the overriding reason for the asymmetric, non-linear relationship between satisfaction and share of wallet is not the absolute level of satisfaction per se. Rather at some point higher/lower levels of satisfaction correspond to a shift in a customer's preference ranking for a brand *vis-à-vis* competitive brands that the customer also uses. As a result, Hofmeyr and colleagues (Hofmeyr *et al.*, 2008; Hofmeyr and Parton 2010) argue that the focus of satisfaction research should shift from absolute satisfaction levels to the relative preference rank that a brand's satisfaction level represents among competing brands used by customers to improve the strength of the relationship between satisfaction and share of wallet.

Relative measures

There is a large body of research confirming the influence of competitive comparisons on both choice and post-purchase evaluations (e.g. Gardial *et al.*, 1994; Rust *et al.*, 2000; Woodruff *et al.*, 1983). For a review of the psychology literature associated with relative thinking in the pre- and post-purchase consumption process, we refer the reader to Keiningham *et al.* (2014).

Relative thinking is central to the consumer decision process. For example, Jacoby and Chesnut (1978, p. 88) argue that "brand loyalty is a function of decision making, evaluative processes. It reflects a purchase decision in which the various brands have been psychologically (perhaps even physically) compared and evaluated on certain internalized criteria, the outcome of this evaluation being that one or more brands was (were) selected."

Similarly, Dick and Basu (1994, pp. 100-101) observe, "Attitudes have been related to behaviors, although it is important to note that one may hold a favorable attitude toward a brand but not purchase it over multiple occasions because of comparable or greater attitudinal extremity toward other brands. For purposes of predictive validity, it is hence advantageous to compare brands that are viewed by consumers to be relevant in a given consumption context. The nature of relative attitudes is likely to provide a stronger indication of repeat patronage than the attitude toward a brand determined in isolation."

Despite this recognition, academic research has overwhelmingly focused on absolute metrics. There are, however, some notable exceptions. Table I provides a brief summary of the research to date regarding the use of relative measures in the scientific literature.

An examination of the research in Table I supports the superiority of relative metrics in linking to customer intentions and behaviors. Interestingly, none of the methods used by these researchers have been widely employed in other scientific investigations. Furthermore, these methods are rarely used by managers.

Instead, the most prominent voices for the use of relative measures in the prediction of share (specifically market share and share of wallet) and the most widely used methodologies come from practitioners. The first widely adopted approach was customer value analysis (CVA), advocated by Bradley Gale (1994) in the book *Managing Customer Value*. One of the primary points of differentiation of the CVA approach was its incorporation of relative brand position in linking customer perceptions to business outcomes, most notably market share. At one time this metric was widely used in industry, although it has fallen out of favor because of underlying statistical issues with the ratios used in the process (Keiningham and Vavra, 2001, pp. 41-44) and the inability of many firms to validate the claimed link to market share (Keiningham *et al.*, 2008).

Hofmeyr *et al.* (2008) introduced a new brand "attitudinal equity" (AE) measure using the Zipf distribution (Zipf, 1935)[1]. The AE measure was calculated by transforming satisfaction (or other perceptual/attitudinal metrics) into relative ranks. Specifically, to transform a customer's satisfaction ratings to ranks, the highest satisfaction rating a customer gave to a brand in his/her usage set would be assigned a "1," the second highest a "2," and so on; in the case of ties, the average is used for the ranks that would have been used had there been no ties. These ranks were then

Table I.
Summary of the research to date regarding the use of relative measures

Study	Setting	Study type	Relative metric operationalization	Outcome	Most important findings/ propositions
Wind (1970)	Electronics industry	Research paper	Two relative metrics are used: 1. Relative attitude towards an ideal supplier 2. Relative attitude towards competitors (i.e. second favorite supplier)	Share-of-wallet	The relative attitude towards competitors is found to be one of the most important indicators of source loyalty
Hauser (1991)	Major consumer-product category	Research paper	Satisfaction rating relative to competition	Primary brand share	Relative scales are found to significantly outperform absolute scales in linking to the primary brand share
Hardie *et al.* (1993)	Retailing industry (Orange juice purchases)	Research paper	Econometric reference-dependent choice model (multinomial logit formulation)	Brand choice	Reference dependent models clearly outperform nonreference-based models, resulting in a better prediction of brand choice
Dick and Basu (1994)	na	Conceptual Paper	Relative attitude defined as the degree to which a customer's evaluation of one product/brand dominates that of other alternatives	Repeat patronage	The inclusion of relative attitudes is likely to result in higher predictive ability for loyalty compared to single-brand attitudes
Van den Putte *et al.* (1996)	1. Broadcasting industry 2. National/regional elections	Research paper	Two relative scales are used: 1. Indirect relative rank order scale 2. Direct relative rank order scale	1. Buying intention 2. Voting intention	Behavioral alternative models applying direct relative rank order scales have the best predictive power, significantly improving average explained variance of behavioral intentions compared to standard, non-relative scales

(*continued*)

Table I.

Study	Setting	Study type	Relative metric operationalization	Outcome	Most important findings/ propositions
Varki and Rust (1997)	Financial services industry	Research paper	Refinement of analysis of variance (ANOVA) for attribute satisfaction ratings	Customer satisfaction	The refined ANOVA-method allows firms to identify their relative performance to competitors at an attribute level, allowing for a better management practice
Bolton *et al.* (2000)	Financial services industry	Research paper	Gain/loss satisfaction scores by comparing focal brand and competitor ratings	Repeat patronage	Customers make re-patronage decisions on the basis of prior re-patronage intentions or behavior, updated by comparing their prior satisfaction level with the company vs that with the competitor(s)
Olsen (2002)	Retailing industry	Research paper	Comparative-attribute based survey format (i.e. quality/ satisfaction questions for different alternatives are posed in sequence, making them salient for comparative evaluation)	Repurchase frequency	Using a comparative assessment, as opposed to an absolute measurement, results in higher predictive power and stronger relationships between quality, satisfaction and loyalty
Kumar (2002)	IT products and services industry	Research paper	Satisfaction gains and losses are computed using the proportional difference between the focal and competing firms	Repurchase intentions	Customers' repurchase intentions depend both on the satisfaction level with the supplier in question and the corresponding satisfaction level and costs of its referent competitor

(*continued*)

Study	Setting	Study type	Relative metric operationalization	Outcome	Most important findings/ propositions
Bowman and Narayandas (2004)	Metal industry	Research paper	Satisfaction with the closest competitor (0 if lower than focal vendor; 1 if equal or higher than focal vendor)	Share-of-wallet	Satisfaction with the closest competitor has a direct, negative impact on share-of-wallet
Rust *et al.* (2004)	Airline industry Electronic stores Facial tissues Grocery stores Rental cars industry	Research paper	Customer ratings on several customer-equity drivers are collected for four to five leading brands in each industry, and imputed in a multinomial logit regression model	Customer lifetime value	The developed CLV-model allows considering the impact of competitive responses on a firm's customer equity, and provides insight into competitive strengths and weaknesses
Ahearne *et al.* (2007)	Pharmaceutical industry	Research paper	Average ratings of competition are subtracted from the focal vendor's service quality and relationship quality measures	Share-of-wallet	Relative service quality evaluations are found to drive relationship quality, which in turn affects share-of-wallet

Table I.

transformed to share of wallet estimates using the Zipf distribution. The parameters[2] of the Zipf distribution were determined by fitting the relationship between the rank of a brand and the corresponding share of wallet that the customer allocated to that brand. (For the remainder of this paper, we will refer to this model as Zipf-AE.)

The results of the Zipf-AE approach showed a large improvement in model R^2. In particular, Hofmeyr *et al.* report that the average R^2 between customer satisfaction and customers' share of wallet using absolute measures was 0.24, while using the rank-based Zipf-AE transformation resulted in a 0.44 R^2.

Keiningham *et al.* (2011) introduced a power law for transforming relative "ranked" satisfaction into share of wallet predictions which they called the Wallet Allocation Rule (WAR). Satisfaction ranks were calculated using the same approach as Hofmeyr *et al.* (2008). WAR is a fixed parameter model; as such, no estimation (i.e. data fitting) is required to estimate the relationship between rank transformed satisfaction and share of wallet. Keiningham *et al.* (2011) report that changes in customers' WAR scores and changes in their share of wallet over time showed a correlation of approximately 0.8, which corresponds to an R^2 of approximately 0.6.

Recently Louw and Hofmeyr (2012) proposed what they described as "an improvement to the original measure of brand attitudinal equity proposed by Hofmeyr *et al.* (2008, p. 10)" which they refer to as a measure of "power of the mind" (PM). As with Hofmeyr *et al.* (2008), the calculation of PM is also based upon the Zipf distribution. (For the remainder of the paper, we will refer to this model as Zipf-PM.)

The primary distinguishing characteristic between the Zipf-AE and Zipf-PM approaches is that Zipf-PM uses "the share that a brand's rating achieves as a percent of the sum of a respondent's ratings of relevant brands" in the Zipf distribution equation (Louw and Hofmeyr, 2012, p. 11).

Louw and Hofmeyr (2012) report that the Zip-PM approach has a higher correlation to share of wallet "by a very small margin" (p. 14) than the Zip-AE and WAR approaches. It is important to note, however, that the comparison made in their investigation was not apples-to-apples; WAR and Zipf-AE were calculated using a single satisfaction question, whereas Zipf-PM was calculated using a combination of two questions in their comparison. Even with this difference, however, there was very little difference in terms of variance explained between the three approaches.

The Zipf and WAR approaches have received a great deal of attention by market researchers. Moreover, both the Zipf-AE (Hofmeyr *et al.*, 2008) and WAR (Keiningham *et al.*, 2011) approaches have received important industry awards for innovation (Gesulado, 2011; Humphrey, 2008).

The primary use of these approaches in practice is within specific products offered by two of the world's largest market research firms. Specifically, Ipsos and TNS use these power laws as core components of their brand equity and customer experience management approaches. As a result, it would be difficult to overstate their use by managers through the use of products offered by these firms. Even if we assume 100 percent overlap of clients, the research firms using these approaches work with over 5,000 different companies worldwide (Ipsos, 2012).

These approaches are not yet widely used by managers outside of the specific product offerings of these firms. As these approaches are not "black boxes" (i.e. these methods are published) and the creators actively promote these approaches (e.g. Hofmeyr, 2012; Keiningham, 2012), however, marketing managers are increasingly aware of the call for relative metrics to more strongly link satisfaction and share of wallet (e.g. Keiningham *et al.*, 2014).

Moreover, while the call for relative metrics has largely come from practitioners, there is early evidence that the academic community has taken notice. For example, Rust and Huang (2014, p. 4) argue that Keiningham (2014) "show convincingly that relative metrics (relative to competitors) are essential."

Research objectives

The primary purpose of this study is to examine the relationship between relative satisfaction and share of wallet. As noted earlier, the research to date tends to support the superiority of relative perceptual and attitudinal metrics to monadic metrics in correlating to consumer buying behaviors such as share of wallet (e.g. Bowman and Narayandas, 2004; Hofmeyr *et al.*, 2008; Keiningham *et al.*, 2011). Therefore, we hypothesize:

H1. Ranked satisfaction levels are more strongly correlated to share of wallet than are absolute satisfaction levels.

Furthermore, although the empirical research appears to confirm the link between absolute satisfaction and share of wallet across various industries such as fleet trucking (Perkins-Munn *et al.*, 2005), pharmaceutical (Perkins-Munn *et al.*, 2005), institutional securities (Keiningham *et al.*, 2005), retail banking (Baumann *et al.*, 2005), processed metals (Bowman and Narayandas, 2004), and grocery retailing (Mägi, 2003; Silvestro and Cross, 2000), the majority of this research has relied on cross sectional data. Although longitudinal examinations of the effect of customer satisfaction on other performance measures have found a positive relationship to customer retention (Bolton, 1998), firm revenues and shareholder value (Anderson *et al.*, 2004), the impact on share of wallet is limited. One exception is the longitudinal share of wallet study by Cooil *et al.* (2007) where results indicate a positive relationship between changes in satisfaction and changes in share of wallet over time. In line with these findings, we would expect longitudinal ranked satisfaction levels to link to changes in share or wallet over time. Therefore we hypothesize:

H2. Changes over time in ranked satisfaction levels are more strongly correlated to contemporaneous changes in share of wallet than are changes in absolute satisfaction levels.

In addition to testing the two hypotheses above, another important goal of this investigation is to provide insight into the most widely used approaches for linking satisfaction and SOW in practice, i.e., WAR (Keiningham *et al.*, 2011), Zipf-AE (Hofmeyr *et al.*, 2008), and Zipf-PM (Louw and Hofmeyr, 2012). In particular, we examine each of the proposed power laws to determine their efficacy in predicting SOW from ranked satisfaction. As noted earlier, to date there is no research in the peer-reviewed scientific literature that examines these various methods to determine their efficacy. Also, we seek to identify better approaches (if any) to link relative satisfaction levels to share of wallet.

Data and measures

Data collection

The data were collected by a large marketing research firm as part of its global norms database. In total, the data consisted of 79,543 customers providing 258,743 observations regarding the brands that they use within a particular industry category. Each respondent in the database used two or more brands in the category (i.e. single-brand users were not included in our database for analysis since their SOW is, by definition, one).

Industries and brands. Data included ratings of over 650 brands in 20 industries. Airlines represented the largest industry in terms of number of respondents, although it should be noted that retail was broken out into more homogeneous subgroups. The complete industry breakdown is: airline (44.9 percent), asthma Rx OTC (0.4 percent), automobiles (0.3 percent), baby retail (1.8 percent), beauty (1.7 percent), clothing retail (2.4 percent), credit card (4.3 percent), DIY retail (0.7 percent), drugstores (1.0 percent), electronics retail (2.0 percent), furniture (2.9 percent), general retail (8.0 percent), grocery retail (13.9 percent), mass merchandise retail (0.5 percent), mobile phone carrier (0.03 percent), office supply (0.6 percent), personal computers (0.2 percent), pharmacy (1.6 percent), printer supplies (2.1 percent), and retail banking (10.7 percent).

Countries. Respondents were sampled from 15 countries, with the majority from the USA. The percentage of respondents from each country is: Australia (0.4 percent), Brazil (3.3 percent), China (0.8 percent), Denmark (0.6 percent), Finland (0.5 percent), Germany (0.6 percent), Italy (8.2 percent), the Netherlands (0.4 percent), Norway (0.6 percent), Peru (0.3 percent), South Africa (0.2 percent), Sweden (0.6 percent), Turkey (1.1 percent), the UK (10.8 percent) and the USA (71.7 percent).

Gender. In terms of total respondents, 51 percent of respondents are female, 49 percent male. The percentage of female respondents for each country is: Australia (30 percent), Brazil (43 percent), Denmark (31 percent), Finland (40 percent), Italy (29 percent), the Netherlands (48 percent), Norway (34 percent), Peru (31 percent), South Africa (20 percent), Sweden (35 percent), Turkey (21 percent), the UK (52 percent), and the USA (53 percent). Gender was not available in the database for Chinese and German respondents.

Age. The average age for all respondents is 49. The average age for respondents in each country is: Australia (48), Brazil (40), China (34), Denmark (49), Finland (45), Germany (38), Italy (48), the Netherlands (47), Norway (45), Peru (41), South Africa (47), Sweden (49), Turkey (34), the UK (48) and the USA (50).

Longitudinal data. A subset of these respondents (all from the USA) were contacted 6 months following the initial survey to provide longitudinal information regarding changes in satisfaction ratings and changes in share of wallet. The longitudinal data consisted of 1,138 customers providing 2,686 observations on the same brands in both periods 1 and 2. These customers provided a total of 3,228 rankings in period 1 and 3,365 rankings in period 2. These 1,138 customers were chosen because they ranked at least two brands in each period. We needed at least two brands from each customer in period 1 in order to be able to use their information to help estimate model parameters. Also, we needed at least two brands per customer in period 2 in order to estimate SOW $<$ 100 percent (i.e. when number of brands equal one, SOW is by default 100 percent).

Gender distribution for the longitudinal sample is approximately even (51 percent male, 49 percent female) with an average age of 55.6. Breakdown of respondents by industry is as follows: grocery (13.4 percent), drugstore (13.4 percent), pharmacy (4.2 percent), mass merchandisers (10.1 percent), retail bank (0.5 percent), asthma Rx (7.9 percent), DIY (17.0 percent), office supply (13.5 percent), airline (12.2 percent), computers (3.1 percent), mobile phone carrier (0.4 percent), and automobiles (4.1 percent).

Constructs and measures

Customer satisfaction. Following Mittal *et al.* (1999) we measured overall satisfaction with the brand using a single item (1 = completely dissatisfied, 10 = completely satisfied). Satisfaction levels were converted to ranks using the approach of Hofmeyr *et al.* (2008) discussed earlier.

It is important to note that relative "ranked" satisfaction is not a single-item construct in the commonly used sense. Rather ranks for customers when "number of brands$\geqslant$2" are based upon consumers' perceptions of multiple brands. In example form, imagine that Brand A has a 7 in satisfaction on a ten-point scale. Its rank will depend on Brand B. If Brand B rates a 5, then Brand A is rank = 1. If Brand B rates a 9, then Brand A is rank = 2. In other words, the same satisfaction level can result in different ranks as information from all brands used by a respondent is used to determine rank. (Note: In this investigation, all respondents used two or more brands.)

With regard to the use of single-item measures in general, although marketing academics typically prefer multi-item measures, single-item measures of overall satisfaction have been used in many prior studies and shown to perform adequately (e.g. Bolton, 1998; Bolton and Lemon, 1999; Cooil *et al.*, 2007; Crosby and Stephens, 1987; Drolet and Morrison, 2001; Mittal and Kamakura, 2001; Mittal *et al.*, 1998, 1999).

Bergkvist and Rossiter (2007) have demonstrated that single-item measures achieve the same predictive ability as multi-item measures, provided that the focal construct is concrete and singular in nature. Satisfaction would appear to meet this standard. Zeithaml *et al.* (2006, p. 170) observe, "Customer satisfaction is the most widely used perceptual metric because it is generic and can be universally gauged for all products and services (including nonprofit and public services). Even without a precise definition of the term, customer satisfaction is clearly understood by respondents, and its meaning is easy to communicate to managers."

Moreover, psychometric analyses conducted by Drolet and Morrison (2001) finds that the incremental information from even the second or third item in a multi-item scale contributes very little to the information obtained from the first item in a multi-item scale. They also find that "added items actually aggravate respondent behavior, inflating across-item error term correlation and undermining respondent reliability" (p. 196).

Of particular relevance to this investigation, Hofmeyr *et al.* (2008) and Keiningham *et al.* (2011) specifically create ranks based upon responses to a single-item measure. This is not surprising given that in practice most firms use single-item measures of satisfaction (Morgan *et al.*, 2005), and these approaches were developed in large part by industry practitioners. Therefore, it is appropriate to apply this same approach in our investigation of these methods.

It is important to note, however, that the longitudinal data examined in this analysis also contained the survey measures used in the ACSI to measure overall customer satisfaction, specifically: overall satisfaction (as used in the single item measure), performance relative to expectations, and performance relative to the customer's ideal (Fornell *et al.*, 1996). Therefore, to be certain that our findings were robust we compared the overall satisfaction measure with two reliable composites of these three questions: both the average of all three and the first principal component of the three items. The average and first principal component are essentially the same (the correlation between the two summaries is 1.000 across both periods) and overall satisfaction has a correlation of 0.95 with each. Given this equivalence, the single-item measure is preferred as the most direct estimate of overall satisfaction.

Share of wallet. Following the approach of Cooil *et al.* (2007), share of wallet was measured as the percent of spending in the category that respondents allocate to the various brands that they use.

Analysis

Description of the relationship between SOW and rank

As noted earlier, research consistently finds that correlation between satisfaction and SOW (at the individual customer level) is very weak. A core argument of the Zipf-AE, Zipf-PM, and WAR approaches under investigation is that relative "ranked" satisfaction is more strongly correlated to SOW. Therefore, the first step was to test the accuracy of this claim.

Table II summarizes the correlations and partial correlations between SOW and both rank and satisfaction (absolute). It also includes correlations with logarithmic transformations of each variable and the logit transformation of SOW[3].

The correlations of SOW, and transformations of SOW, with *Rank* and log(*Rank*) are invariably stronger than the correlations of SOW, and the transformations of SOW, with the two versions of *Satisfaction*. Nevertheless all correlations are highly significant ($p < 0.001$; which is not surprising given the sample size, $n = 258{,}743$). The strongest relationships are for log(*Rank*) with SOW and logit(*SOW*); log(*Rank*) explains 30 percent (or $r^2 \times 100\%$, with $r = -0.545$) of the variance in SOW and 29 percent of the variance of logit(SOW) ($r = -0.536$). The largest nominal correlation with *Satisfaction* are for *Satisfaction* with the logit(SOW) ($r = 0.239$), which indicates it accounts for 5.7 percent of the variation in logit(SOW).

Remarkably, the correlations of *Rank* and log(*Rank*) with SOW and its transformations still remain strong and quite significant when we condition on *Satisfaction* levels, as seen from the partial correlations (the percent variance explained ranges from 19 to 26 percent in each case). In contrast, the partial correlations of *Satisfaction* and log(*Satisfaction*) with SOW and its transformations are actually negative, and correspond to R^2 values that are below 1 percent in absolute value in every case.

Our results provide strong evidence of the superiority of relative ranked satisfaction to absolute satisfaction in linking to SOW.

Investigating the models[4]

The next step in our analysis was to investigate the efficacy of the three most widely used power laws (i.e. Zipf-AE, Zip-PM, and WAR) in predicting SOW. A fair assessment, however, requires that we compare these power laws to other models that would be reasonably expected to perform similarly based upon the properties of these models.

Table II. Correlations between transforms of SOW and transforms of *Rank* and *Satisfaction*

					Partial correlations after removing *Satisfaction*		Partial correlations after removing rank (as log(*Rank*))	
	Rank	Log (*Rank*)	*Satisfaction*	Log (*Satisfaction*)	*Rank*	Log (*Rank*)	*Satisfaction*	Log (*Satisfaction*)
SOW	−0.484	−0.545	0.237	0.192	−0.437	−0.505	−0.010	−0.027
Log(SOW)	−0.492	−0.521	0.231	0.191	−0.448	−0.479	−0.003*	−0.016
Logit(SOW)	−0.491	−0.536	0.239	0.195	−0.443	−0.494	−0.002**	−0.019

Notes: $n = 258{,}743$. Except as indicated, all correlations are significant at the $p < 0.001$ level; $^*p = 0.124$; $^{**}p = 0.209$

The Zipf functions imply a Pareto decay in SOW as rank increases, which is distinct from a geometric decay and more rapid than the linear decay of the WAR model (when there are more than two brands). Therefore, to provide an additional reasonable point of comparison for the Zipf-AE and Zipf-PM power laws, we examine the effectiveness of the truncated geometric model in using ranked satisfaction to predict SOW.

Whenever possible, we examine three versions of these discrete distributions: a fixed-parameter version, a one-parameter version (i.e. the parameter does not vary by the total number of brands), and what we label as a nine-parameter version (i.e. the parameter varies by the total number of brands used; we consider customers who use from two to ten brands). It is important to note that there is no one-parameter version of WAR, and no established fixed-parameter version of the truncated geometric. In total, we explore ten versions of the discrete distribution models by including fixed-parameter, one-parameter and nine-parameter versions of the various models.

Additionally, because hierarchical regression models are commonly used in research and practice to assess the relationship between satisfaction and SOW (e.g. Keiningham *et al.*, 2003) we investigated these models as a point of comparison. In each of these models, a random effect at the customer level is used to accommodate the dependence among observations from the same customer within a product category. Specifically, we consider four hierarchical regression models (where for each of set of predictors, we estimate one version with common parameters across all m-categories, where m represents the total number of brands, and another with separate parameters within each m-category).

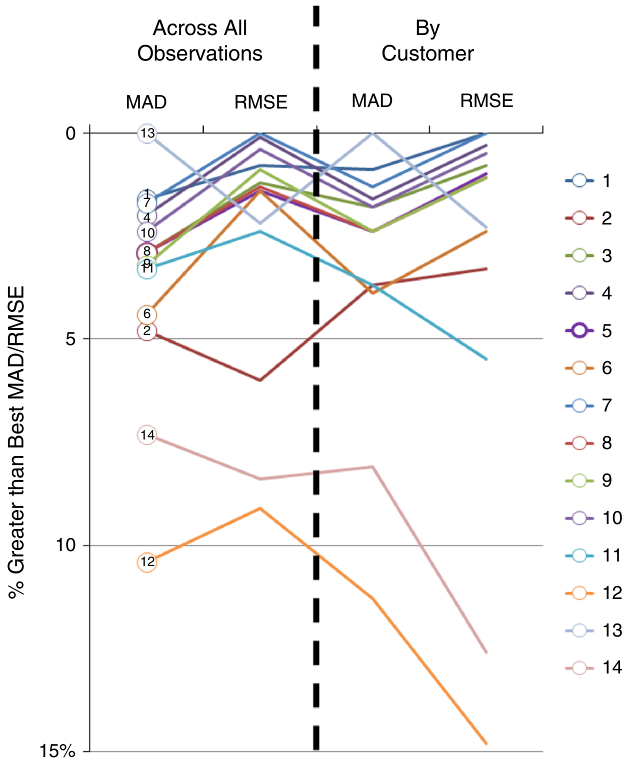
Overall model performance (cross-sectional)

To evaluate the overall performance, we first examined each model's ability to link customer satisfaction (absolute or relative ranked satisfaction) with SOW for the same time period. We assess each model's performance in four ways: mean absolute deviation (MAD), and root mean squared error (RMSE) across all observations and also by customer. Figure 1 shows the performance of each of the models relative to the best performing models[5].

The fixed-parameter versions of the discrete distribution models do remarkably well overall. Among these distributions, the fixed-parameter Zipf-AE model is best in terms of MAD, both overall and per customer, and it actually outperforms all models (including the regression models) in terms of average customer RMSE. The nine-parameter version of Zipf-AE is the best performer in terms of overall RMSE. Nevertheless, the discrete distributions generally do quite well: eight of the other ten discrete distributions have RMSE values that are within 1.5 percent of the best fit. The one exception is the fixed parameter Zipf-PM which has an RMSE that is 6 percent larger overall, relative to the best performing nine-parameter Zipf-AE model.

The nine-class regression with log(*Rank*) is actually the best performing model in terms of MAD, and it is just ahead of the fixed parameter Zipf-AE with MAD values that are 1.6 percent and 0.9 percent larger overall, and per customer, respectively. This regression model is also uniformly the best among the four regression alternatives, but paradoxically it does not fit as well in terms of RMSE, where it actually achieves the 10th and 9th highest overall and per customer RMSE, respectively. Still, even in these cases its error rates are only larger than the lowest RMSE values by 2.2v overall, and 2.3 percent per customer.

By contrast, the regression models based on *Satisfaction* are uniformly the worst models in every case, and here the error rates are substantially larger than the best



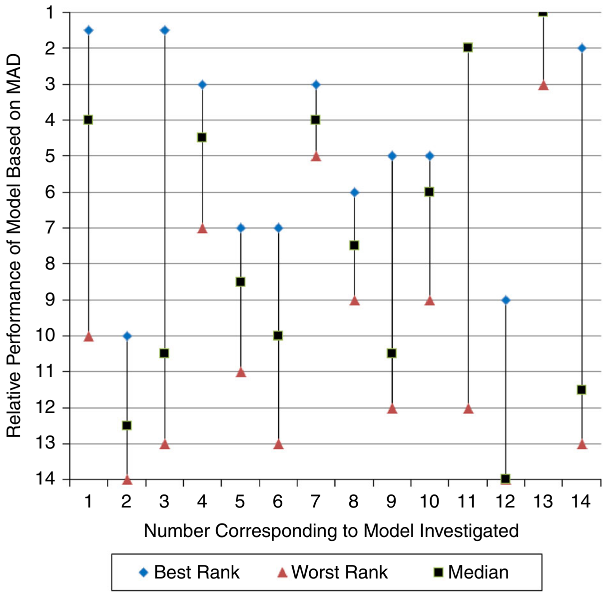
1. Zipf-AE, Fixed Parameter
2. Zipf-PM, Fixed Parameter
3. WAR, Fixed Parameter
4. Zipf-AE, One-Parameter
5. Zipf-PM, One-Parameter
6. Truncated Geometric, One-Parameter
7. Zipf-AE, Nine-Parameter
8. Zipf-PM, Nine-Parameter
9. WAR, Nine-Parameter
10. Truncated Geometric, Nine-Parameter
11. Logit(SOW_{ij}) = $\beta_0 + \beta_1$Log(*Rank*)+ β_2Log(*Total Brands*+1−*Rank*), 1 Class
12. Logit(SOW_{ij}) = $\beta_0 + \beta_1$*Satisfaction*+β_2(*Total Brands*), 1 Class
13. Logit(SOW_{ij}) = $\beta_0 + \beta_1$Log(*Rank*)+ β_2Log(*Total Brands*+1−*Rank*), 9 Classes
14. Logit(SOW_{ij}) =$\beta_0 + \beta_1$*Satisfaction*, 9 Classes

Figure 1.
Model performance overall and at the customer level in terms of mean absolute deviation (MAD) and root mean squared error (RMSE)

model in every instance. Although the nine-class version of this model is the better performer, even its error rates range from being higher by 7.3 percent (MAD overall) to 12.6 percent (RMSE per customer).

Overall model performance by number of brands used. In addition to examining overall performance, we investigated whether the number of brands used by the customer affect which model performs best. Figure 2 provides a comparison of model performance by the number of total brands that are used by the customer. An examination of Figure 2 shows that the relative performance of most models varies widely depending upon the number of brands used by the customer[6].

The fixed-parameter versions of Zipf-AE and WAR are the best in the two-product category with MAD values of 20.5 percent. WAR and Zipf-AE are equivalent in this case.



1. Zipf-AE, Fixed Parameter
2. Zipf-PM, Fixed Parameter
3. WAR, Fixed Parameter
4. Zipf-AE, One-Parameter
5. Zipf-PM, One-Parameter
6. Truncated Geometric, One-Parameter
7. Zipf-AE, Nine-Parameter
8. Zipf-PM, Nine-Parameter
9. WAR, Nine-Parameter
10. Truncated Geometric, Nine-Parameter
11. Logit(SOW_{ij}) = $\beta_0 + \beta_1$Log($Rank$)+ β_2Log($Total\ Brands$ + 1−$Rank$), 1 Class
12. Logit(SOW_{ij}) = $\beta_0 + \beta_1$$Satisfaction$+$\beta_2$($Total\ Brands$), 1 Class
13. Logit(SOW_{ij}) = $\beta_0 + \beta_1$Log($Rank$)+ β_2Log($Total\ Brands$ + 1−$Rank$), 9 Classes
14. Logit(SOW_{ij}) = $\beta_0 + \beta_1$$Satisfaction$, 9 Classes

Figure 2.
Relative model performance in terms of total brands used based upon mean absolute deviation (MAD)

This is the only category where the nine-class regression with log(*Rank*) is not the best model, and even in the two-category case this regression model is nearly the best with a MAD that is 20.6 percent (relative to the best MAD of 20.5 percent).

The nine-parameter Zipf-AE model and the nine-class regression with log(*Rank*) are the best overall performers across categories, and the Zipf-AE models are always among the top 5 models when total brands is less than seven ($m \leqslant 6$). Finally the regression models based on *Satisfaction* are the worst models overall, in terms of median rank across categories, although the nine-class regression on *Satisfaction* is the second best model in the last category ($7 \leqslant m \leqslant 10$). The regression models based on *Satisfaction* are uniformly the poorest performers when there are four or fewer total brands ($m \leqslant 4$).

It is important to note that while the relative performance of most models varies by the number of brands used, MAD values decrease as the total number of brands used increases (see Figure 3), which is to be expected, given that one is predicting smaller SOW values as the total number of brands increases. Across models, thelowest MAD values decrease by 64 percent as total brands increase across the six categories, and it ranges from 20.5 percent (when $m = 2$) to 7.4 percent (when $7 \leqslant m \leqslant 10$).

Table AIII.

Change in SOW		Change in Logit(SOW)	
Rank	−0.285	Rank	−0.278
Log(*Rank*)	−0.332	Log(*Rank*)	−0.328
Satisfaction	0.066	*Satisfaction*	0.111
Log(*Satisfaction*)	0.078	Log(*Satisfaction*)	0.098
Log((*Total Brands*)+1–*Rank*)	−0.038[a]	Log((*Total Brands*)+1–*Rank*)	−0.010[a]
Logit(*Rank*/(*Total Brands*+1))	−0.133	Logit(*Rank*/(*Total Brands*+1))	−0.180
Logit(*Satisfaction*/(Maximum(*Satisfaction*)+1))	0.069	Logit(*Satisfaction*/(Maximum(*Satisfaction*)+1))	0.112
Recommend Intention	0.065	*Recommend Intention*	0.070
Log(*Recommend Intention*)	0.056[a]	Log(*Recommend Intention*)	0.061[a]
Net Promoter	0.067	*Net Promoter*	0.073
Log(*Net Promoter*)	0.068	Log(*Net Promoter*)	0.073

Notes: $N \geqslant 2,686$ for each correlation. [a]All correlations are significant at the level $p < 0.001$, except for the correlations of change in log((*Total Brands*)+1–*Rank*) with change in SOW and Change in Logit(SOW)(where $p = 0.050$, and $p = 0.608$, respectively), and the correlations of change in log(Recommend Intention) with change in SOW and change in logit(SOW) (where $p = 0.004$, and $p = 0.001$, respectively)