FAC-SEM: A Methodology for Modeling Factorial Structural Equations Models, Applied to Cross-Cultural and Cross-Industry Drivers of Customer Evaluations

Dawn Iacobucci
Northwestern University

Doug Grisaffe
Walker Information

Adam Duhachek
Northwestern University

Alberto Marcati
Luiss Guido Carli University, Italy

This article presents factorial structural equations models (FAC-SEM). An experimenter conducting ANOVAs on means can use FAC-SEM on covariance matrices. To illustrate, the authors model customer evaluations and examine how construct relationships vary by sector and country. The methodological approach presented in this article is new to the literature and easily implemented.

Keywords: *customer satisfaction; structural equations models; international data; cross-cultural data analysis*

This article is primarily a methodological investigation. We hope statisticians and methodologists will find the techniques we present interesting and useful. But most techniques are not created to be goals in and of themselves; rather, they are usually created to be analytical means to other goals, typically substantive theoretical investigations. We hope that readers interested in a variety of substantive areas will also find something of value in this article. Specifically, the theoretical content to which the analyses are applied in this article involves three elements: (a) customer evaluations, (b) cross-cultural comparisons, and (c) cross-industry comparisons. Indeed, any one particular reader may wish to focus only on the section(s) that speak most directly to their research needs and skim or skip the other sections of this admittedly complex article. By way of an overview, let us describe the methodology abstractly and the three substantive elements to which we just alluded.

Journal of Service Research, Volume 6, No. 1, August 2003 3-23
DOI: 10.1177/1094670503254271
© 2003 Sage Publications

First, the methodology involves a multigroup approach to structural equations modeling (i.e., lisrels). The multiple groups will have derived from a factorial design, common to many behavioral experiments, perhaps more rare in survey design. Factor A is varied, crossed with Factor B, and their main effects are examined, as is their interaction. In analysis of variance, those effects involve comparisons of means. In our methodology, the comparisons will involve covariance matrices or, more to the point, the structural equations models (SEM) used to fit the covariance matrices that arise from the $A \times B$ factorial design. Instead of an A_iB_j cell containing a mean, our cells contain covariance matrices that describe relationships between constructs and variables. Let us turn now to describe those interrelated constructs as well as the Factors A and B.

The constructs that we will be exploring include components of customer evaluations. Numerous business-to-business customers were surveyed about their perceptions of a purchase regarding its quality, value, likelihood to repeat purchase, and so on. If we had had only one sample, the article would be very straightforward. An SEM would be posited, the hypothetical links supported and derived from the literature, the covariance matrix estimated, the model fit and tested, and the results interpreted: "Does value contribute to repeat purchasing?" and so on.

The point of this article is to examine modeling approaches for data that are more complicated. We do not have a single sample; rather, we have a number of covariance matrices that may be organized via a factorial design. That design is composed of two factors. Factor A is the cross-cultural factor, or the country from which the data were obtained. We have many, many countries and cultures represented in our data, and this feature of our data helps to make this investigation richer than the normal (usually two-country) cross-cultural investigation. Factor B is the sector from which the business-to-business customer purchased—either a "good" or a "service." Given that the data are "real" data, the content is somewhat disguised, but the good was a high-tech good, and the service was a financial service. Together, our factorial structural models will test whether the drivers of positive customer evaluations vary as a function of the two factors: the product (good or service) being evaluated, the cultural desires of the customer base making the evaluations, and their possible contingent interaction.

That description might have been a bit abstract. More specific examples of the class of research and business questions that may be addressed using our models follow. Our methods will allow researchers and managers to answer questions within a culture and sector, such as,

- Are my (financial service) customers in Singapore more sensitive to the quality of sales representatives than to the responsiveness of the customer service?

The models also enable hypothesis testing within a culture across sectors, such as,

- Is customer satisfaction among the French driven by value for the purchase of a good but sales reps for the purchase of a service?

The models address questions that compare across cultures, perhaps,

- Do German customers care more about the quality of the core good being purchased than do Dutch customers?

And finally, the models allow testing interaction premises in which culture and sector vary, for example,

- Do Brazilians care more about price on computers than the Japanese care about customer representatives' responsiveness for the credit cards they carry?

All of these kinds of questions are of keen interest to the marketing manager responsible for a global multinational or for investigations into markets to enter or from which to withdraw. Managers making allocation decisions should also find these questions, and therefore the methods, pertinent.

To date, cultural comparisons have been hampered by inadequate technologies or limited perspectives, for example, a focus on mean differences (such as through analysis of variance) usually attributed to cultural differences in response styles. Our methods will allow the focus on relationships between constructs, having equated measurement differences.

In this article, we try to present a thorough case on all three of these substantive theoretical elements: customer

We are grateful to *JSR* editor Roland Rust and three anonymous reviewers; Richard Bagozzi, Bridgette Sullivan-Taylor, and participants in the MAXX European seminar; Richard Oliver and colloquia participants in the marketing department at Vanderbilt University, the University of Maryland, and New York University; Charles Hulin and the quantitative psychology department at the University of Illinois; Bruce Spencer of the Statistics Department at Northwestern University; and Alex Chernev, Rob Kozinets, Angela Lee, Christie Nordhielm, John Sherry, Brian Sternthal, and Alice Tybout for their insightful comments regarding this research and article. Please address correspondence to the first author at Department of Marketing, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, phone: 847-491-2722; fax: 847-491-2498; e-mail: d-iacobucci@kellogg.northwestern.edu.

evaluations, services and goods sectors, and cultural differences; but it is important to note that the modeling techniques are more general still than their applicability to cross-cultural and cross-industry studies. For example, factorial experimental designs abound, and usually researchers are interested in testing hypotheses about mean differences. Those tests are fine, and we are not suggesting they be replaced with the techniques in this article. However, the techniques in this article would allow supplementing such investigations. Usually, the researcher conducting a laboratory or field factorial experiment collects data on multiple measures and may well have hypotheses about how the variables and constructs are interrelated, hypotheses about structure above and beyond the hypotheses about means tested via analysis of variance. The connections among variables and constructs arising in the factorial design may be tested easily using our models.

We tried to keep each section of this article as succinct as possible, but we are trying to cover a methodology as well as three substantive components (customer evaluations, cross-cultural observations, and cross-sector comparisons). To help the reader navigate this article, we note that the organization of this presentation is as follows.

First, we present the customer evaluation model that interconnects the constructs of value, quality, and so on, as per the current status in the marketing literature, supporting the links we propose with some relevant cites. Second, we derive from the cross-cultural literature the marketing distinctions that should be relevant when modeling such customer evaluations. In the cross-cultural presentation, we also touch on some of the typical methodological problems faced when exploring such data to help highlight the advantages of a technique like the one we subsequently present. Third, we describe our data. We describe the specific constructs, variables, and measures as well as the particular countries and cultures involved. Fourth, we turn to the technical part of the article. We present the known models from the literature regarding multigroup approaches for factor analysis. We then briefly remind the reader about the basics of SEM, mostly to establish the notation. We then put the two together, offering our multigroup approach to SEM for factorially structured data. The statistical inferential methodology allows the simultaneous modeling of two factors, in this specific case, the industry and cross-cultural comparisons. This section of the article contains many technical hypotheses about the parameters that may vary or be constrained within a culture or sector or both, that is, either main effect or the interaction. Fifth, we take the new technology and apply it to our data. Sixth and finally, we compare these results to the literature on customer evaluations models to bring the paper round full-circle.

CUSTOMER EVALUATION MODEL AND LITERATURE

The model of customer evaluations in Figure 1 is made up of links derived from the marketing literature. Customer satisfaction ratings on items tapping constructs of attributes (cost, service, product, sales representatives) are conceptualized to operate through mediators (value, quality and easy to do business with the company), which in turn are used to predict the dependent variable of the likelihood with which the customer expects to continue doing business with the company. The reasoning and the literature from which the links among these constructs were deduced are presented here.

Value is defined as a comparison of inputs (costs) to outputs (quality, e.g., Zeithaml 1988), so there should exist links:

Hypothesis 1: cost → value. Greater perceptions of cost should be related to lower perceived value.
Hypothesis 2: quality → value. Greater perceptions of quality should enhance perceptions of value.

The concept "easy to do business with" may be considered the obverse of a cost; thus, there should be a link:

Hypothesis 3: easy → value. If a partner is thought to be so compatible, the cost of doing business with that partner is reduced; hence, ratings of value should be more positive.

Cost is often considered to be a cue to quality used by customers (cf. Voss, Parasuraman, and Grewal 1998); hence, we may test

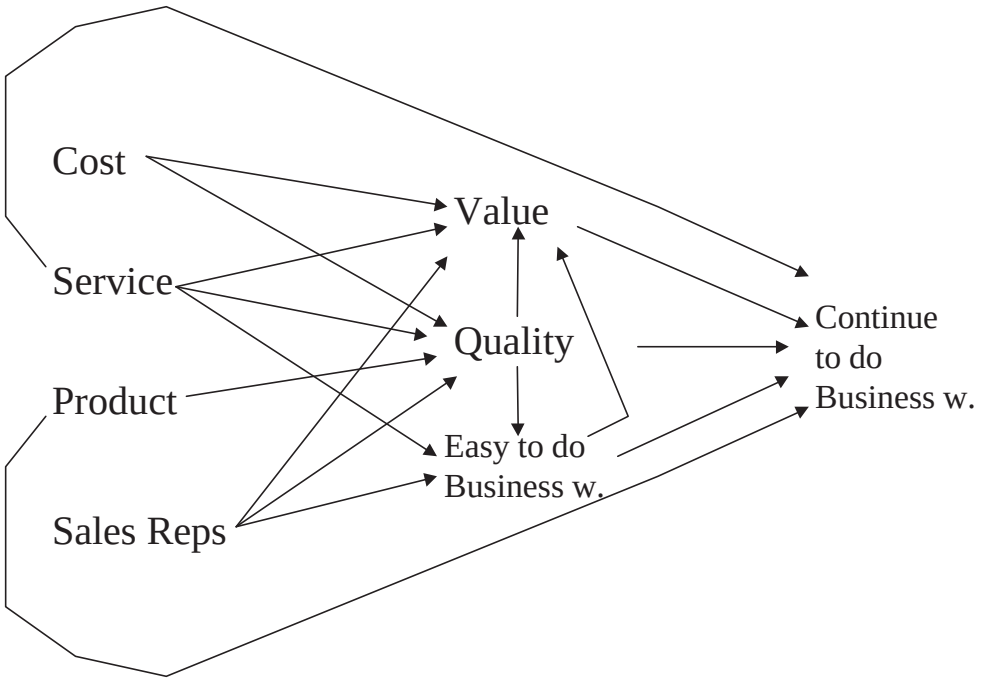
Hypothesis 4: cost → quality. If cost is a cue to quality, then customers who rate their purchases as more costly should also rate their purchases as being of higher quality.

Quality is posited as a multiattribute function (e.g., Boulding et al. 1993; Furrer, Liu, and Sudharshan 2000; Oliver 1997), so links between attributes and quality should be significant:

Hypothesis 5: service → quality;
Hypothesis 6: product → quality; and
Hypothesis 7: sales reps → quality. That is, the better the service, product, or sales representatives, the more favorable should be perceptions of overall quality.

The services marketing literature distinguishes the "core" of one's purchase, which might be a good or a service (e.g., a dinner at a restaurant), from the "supplemental" or "value-added" services (e.g., the pleasantness of the

FIGURE 1
Consumer Evaluation Relationship



interactions with the waiter; Lovelock 1999). Services that facilitate business transactions should therefore be appreciated as adding further value. Thus, we test for links between

Hypothesis 8: service → value and
Hypothesis 9: sales representatives → value. A company's orientation to its customers, whether rated via customer service scales or perceptions of the quality of the company's sales reps, should enhance the customer's perception of the value being provided by that company.

The extent to which a customer perceives a company as one that is easy to do business with should be a function of several components. Reliably high quality is often of paramount importance (e.g., Furrer, Liu, and Sudharshan 2000; Parasuraman, Berry, and Zeithaml 1991), so the higher the quality of the purchase, the easier it should be to retain the present business partner. In addition, enhanced customer service (e.g., a computer manufacturer's 800 help line) should facilitate the ease of continuing to do business together. A frontline staff for service provision can also include the sales representatives (e.g., retail, phone, or online persons assisting at point of sale), that is, any personnel likely to have a "moment of truth" contact with the customer. Together these predictions are

Hypothesis 10: quality → easy (to do business with);
Hypothesis 11: service → easy; and
Hypothesis 12: sales representatives → easy. That is, there are several facilitators of a customer's perception of the facility of a business partnership. The quality of the purchase, the supplemental service provision, and the frontline sales rep staff should all contribute to a customer's believing that the company is one with which it is easy to do business.

Finally, there should exist several precursors to continuing to do repeat business. First, it is fairly well documented that quality enhances repeat purchasing (e.g., Oliver 1997; Rust, Zahorik, and Keiningham 1996), so we would expect replication:

Hypothesis 13: quality → continue. More favorable views of a purchase's quality should lead to a propensity for repeat purchasing.

In addition, there seems to be a persistent belief that value should enhance repeat likelihoods (e.g., Reichheld 1993), even in the absence of empirical support (Ander-

son, Fornell, and Lehmann 1994). Given the debate in the literature, the empirical test will examine whether, for these purchases:

Hypothesis 14: value → continue. If a customer experiences a purchase as being of greater value, the customer should be more likely to purchase in that transaction again.

For two different reasons, there may exist a link between the ease of doing business and the likelihood with which repeat is expected:

Hypothesis 15: easy → continue. Perhaps explicable in terms of *inertia*, the purchase is good enough in quality, value, and so on, that cognitive and physical switching costs and search efforts for suitable alternatives are high. Alternatively, perhaps *relationship marketing* drives the connection, that is, a business relationship fundamentally eases the interactions and enhances the likelihood of repeat.

Finally, repeat intentions may be driven to a great extent simply by a good fit between one's needs and the product offering. That is, the customer in the market for a specific good/service evaluating the current version as performing adequately, then other characteristics of the purchase process may be less relevant. Thus, for the goods data, we would expect to see the link in Hypothesis 16 and, for the services data, in Hypothesis 17:

Hypothesis 16: product → continue and
Hypothesis 17: service → continue. Depending on the central item purchased, whether a good or service, to the extent that it enjoys more favorable ratings, the customer should more likely repeat that purchase, all other things being equal.

These hypotheses and links are collected into Figure 1, which serves as the base model. If one's research goal is theoretical generalization, then Figure 1 would be the model to fit in all sectors and cultures, and the question posed would be of the nature, "For which sector/culture is the model insufficient?" Yet practicing marketing researchers and managers, as well as academics interested in contingent relationships and moderating variables, would be motivated to explore the natures of the variants from the model in Figure 1.

Thus, some of the linkages in Figure 1 may be universally empirically supported but others are likely to vary with cultural contexts or depend on the particular marketplace. For example, researchers point to the difficulties in satisfying heterogeneous markets (Anderson, Fornell, and Lehmann 1994) and to the inherent heterogeneity of service provision, including price fluctuations, suggesting the importance of price signaling quality more for services even than for goods (Voss, Parasuraman, and Grewal 1998). If these premises hold true in our data, we would expect there to be a Hypothesis 18: the price → quality link to be stronger for services than in the data on goods. Cross-cultural variants on this model will also be investigated. For example, the collectivistic cultures of Asia and Latin America might show a Hypothesis 19: a stronger, more persistent link between service and repeat purchasing, given the prevalence of the personnel component in service provision (Lovelock 1999) and the value these cultures attach to relationships. Similarly, more collectivistic cultures may need the interpersonal element as a means of assurance, one of the primary dimensions of service quality (Furrer, Liu, and Sudharshan 2000). An analogous argument could be made for a Hypothesis 20: the sales representatives should be more important in these cultures in affecting good outcomes such as perceptions of high quality than in individualistic cultures, where the product itself, or its attributes, for example, cost, should reign supreme over relational interactions. The cross-cultural terms and relationships are explained in the following section.

THE CROSS-CULTURAL LITERATURE

Cross-cultural research, previously a seeming theoretical curiosity under the purview of social sciences, has become more critical and highly practical to guiding foresightful business strategy and daily business tactics both for global multinational corporations and for invisibly bordered internet providers (Craig and Douglas 2000). Cross-cultural business resources exist that prompt marketers to consider issues as diverse as the linguistic implications of brand names, courteous communication patterns, cultural perceptions of time, local preferences for office design, reactions to ethical dilemmas posed, and the attribution of credit following a successful project implementation (Brake, Walker, and Walker 1995; Roth 1995; Usunier 1996; Vandello and Cohen 1999).

There also exists a substantial marketing literature on cross-cultural issues. For example, researchers compare new product development methods in Japan versus the United States (Song and Parry 1997), strategies for obtaining technology licensing when attempting to enter new markets (Kotabe, Sahay, and Aulakh 1996), and ethnocentric distinctions regarding customers of one culture purchasing goods manufactured by another (Klein, Ettenson, and Morris 1998). Many more marketing research articles have been published that examine cross-cultural issues, and they will be discussed later in the article, where pertinent.

Cross-cultural researchers seek to characterize the "values" of a group—those beliefs so fundamentally and deeply held, so widely shared by the individuals making up the cultural group, as to define that group and distinguish it from other groups (Cooper and Denner 1998; Hofstede et al. 1990; Schwartz 1992). Similarity, or "cultural closeness," (Graham et al. 1988) is a function of cultural members' shared belief systems and has been demonstrated to enhance beneficial outcomes, for example, international negotiations exercises (Campbell et al. 1988).

If within-group similarity defines cultural closeness, cross-cultural comparisons seek between-group differences. Theorizing and empirical comparisons usually begin with the benchmark of Hofstede's (1983) four key dimensions of (a) individualism versus collectivism, (b) power distance, (c) masculinity-femininity, and (d) uncertainty avoidance. Individualism is "a concern for oneself as opposed to concern for the collectivity to which one belongs" (Berry et al. 1992, p. 54). Examples of individualistic countries typically cited include the United States, Canada, the United Kingdom, Australia and New Zealand, and the Netherlands, compared with more collectivistic countries such as Colombia, Peru, Mexico, Singapore, and Hong Kong (Hofstede 1983). Cultures with greater power distance are those that are asymmetric and hierarchically structured (e.g., Mexico, Brazil, Colombia, Peru, France, Hong Kong), compared with more egalitarian societies (e.g., Australia and New Zealand, Switzerland, Germany, the United Kingdom). Cultures said to be masculine are those that emphasize assertiveness, competition, and achievement (e.g., Germany, the United Kingdom, Switzerland, Colombia), in contrast to societies that are more relational and nurturing and emphasize quality of life (e.g., Netherlands, Peru, France, Spain). Countries that seek to avoid uncertainty are risk-averse and conservative, preferring stability (e.g., Argentina, France, Peru, Spain) compared with more dynamic, change-seeking states (e.g., the United Kingdom, Hong Kong, Singapore, the United States).[1]

Furrer, Liu, and Sudharshan (2000) explored Hofstede's (1983) distinctions as they relate to the five dimensions of service quality. For example, they found cultures high on power distance had a negative relationship with service reliability and a positive relationship with the quality of the tangibles of the purchase. Uncertainty avoidance was positively related to service responsiveness and negatively related to the tangibles. Mattila (1999) found Western customers (presumably more individualistic, for example) valued personalization and the physical environment (an exemplar of a tangible) more than Asian customers. Consistent with these findings, in our data we will see, for example, that Japan, a country prototypically high on collectivism and the need to avoid uncertainty, shows more significant relationships with sales representatives (whose assurances, relationships, and own corporate cultures may serve as risk reducers) as drivers of perceptions of quality for the good, not the tangible good itself.

Donthu and Yoo (1998) found cultural differences on levels of expectations, a key driver of service quality evaluations. For example, individualism was correlated with high expectations, particularly relating to the interpersonal interactions with the service provider (e.g., empathy, assurance). Lower power distance was associated with high expectations, particularly on dimensions of reliability and responsiveness. These cultural modifiers would be easily tested using extant methods such as analysis of variance. When the cultural predictions become moderators of relationships between constructs, those classes of differences could be tested using the models we offer in this article.

Cultural groups can differ on their stated behavior intentions. For example, individualists are difficult customers—they are less likely to praise superior service and, in reaction to poor service, are more likely to switch, generate negative word of mouth, and complain than are their communal customer counterparts (Liu, Furrer, and Sudharshan 2001).

Cross-cultural generalizations are defended by appealing to the logic of sampling, whereby labeling a country *individualistic* implies that a random sampling of individuals and their behaviors will yield distributions that appear more idiocentric than allocentric (Oyserman, Coon, and Kemmelmeir 2002; Triandis 1994). Explanations and covariates are sought for country differences, for example, as a function of ecological (climate, resources), socioeconomic, political, historical, and religious contexts (Hofstede 1983; Roth 1995; Schwartz 1992; Smith and Schwartz 1997). For example, various measures of economic development, such as per capita gross domestic product (GDP), show reliable positive correlations with individualism and negative relationships with power distance and uncertainty avoidance (Hofsted 1983; Smith and Schwartz 1997). Stronger modeling techniques will further enhance theory development and testing, as we shall demonstrate.

1. Some researchers find different labels more useful, but the constructs are highly similar; for example, Schwartz (1992) distinguished countries along dimensions self-enhancement versus self-transcendence and openness to change versus conservativism, which may capture individualism and risk aversion. Other researchers begin with Hofstede's (1983) fundamental four dimensions and add variants, for example, Confucianism (Nakata and Sivakumar 1996).

METHODS IN CROSS-CULTURAL RESEARCH

Cross-cultural social scientists engage in debate over philosophical methodological issues and the empirical modeling of data. The first philosophical issue derives from whether the researcher believes that general conclusions may be extracted from the analysis of multiple cultures. The emic philosophical view holds that "attitudinal or behavioral phenomena are expressed in a unique way in each culture" (Usunier 1996, p. 142; cf. "relativism" in Berry et al. 1992). This perspective dictates fitting models to data within each observed culture separately, and comparisons are minimized (e.g., Hofstede, Bond, and Luk 1993). In contrast, the etic approach "is primarily concerned with identifying universals" (cf. Berry et al.'s 1992 "absolutism" and "universalism"). Researchers defending this position aggregate respondents across cultures for a "pancultural" analysis (Hofstede, Bond, and Luk 1993; Smith and Schwartz 1997). Both philosophies are defensible on scientific grounds—generalization, abstraction, and parsimony support the etic; whereas contingent interactions and moderators, which enrich and deepen theorizing, support the emic. In the approach that we develop, we include both aggregate, or etic, analyses to allow for the derivation of general principles, as well as disaggregate, emic, analyses conducted per country or per culture to allow for the customization of moderators, as the researchers' needs would dictate.

A second philosophical issue reflects consternation over interpretations of found cultural differences. Cross-cultural studies arise from natural observation and, as a result, theoretical explanations for cultural differences can have confounds (Berry et al. 1992; Lonner and Adamopoulos 1997). This difficulty is exaggerated for studies that contain data on only two cultures; for example, if one were to compare data from the United Kingdom to data on customers of Hong Kong, one might attribute differences to the citizens of the United Kingdom's being individualistic and those of Hong Kong collectivistic, or one might point with equally validity to the United Kingdom's being egalitarian compared with Hong Kong's greater power distance and hierarchical structure. Although obtaining data from two countries seems more than twofold effort than doing so from one, a number of researchers have recognized that obtaining data from even more countries strengthens comparative theory testing (e.g., De Wulf, Odekerken-Schröder, and Iacobucci 2001), such as introducing a country that was not both individualistic and egalitarian would begin to tease apart the nature of a U.K./Hong Kong difference. The data in this project represent many countries, which will enhance our theoretical conclusions because the numerous countries constrain and negate more alternative explanations.

There are also empirical modeling issues of concern to the cross-cultural analyst. Researchers discuss "measurement" issues and "structural" issues. This distinction lends itself nicely to the SEM paradigm discussed subsequently.

Measurement issues are motivated by concerns of "response styles" and come in the form of means and variances (Craig and Douglas 2000, chaps. 10, 11). Greenleaf (1992), Mullen (1995), and Steenkamp and Baumgartner (1998) are nice expositions on the process by which the researcher would test for the equivalence in origins (means) and scalars (standard deviations or measurement units). More positive means tend to come from Asian and Latin American cultures, which are said to value acquiescence and so tend to yield more socially desirable responses as a "courtesy bias" (Usunier 1996; Baumgartner and Steenkamp 2001; also see Donthu and Yoo 1998). Response styles can also differ in terms of variances; for example, respondents from cultures labeled "enthusiastic" (e.g., the United States, France, Italy) tend to select extreme scale values more often than those from cultures expecting more conservative behavior (e.g., Japan, Germany; Smith and Schwartz 1997; Usunier 1996).

Although researchers acknowledge the possibility of these kinds of measurement differences across cultures, their research interests usually focus more on whether there exists equivalence in the structure of the interrelationships between the focal constructs across samples. They reason that an assessment of the equivalence of meaning can be conducted by examining covarying indices. Even if a sample from Japan yields higher means and lower variances than one from the United States, a correlation coefficient is standardized across means and variances and should yield similar patterns if the two cultures converge on shared meaning (Schwartz 1992; Smith and Schwartz 1997).

Thus, cross-cultural research usually focuses on structural analyses, for example, multivariate methods such as factor analysis, multidimensional scaling, or cluster analysis (Hofstede 1983; Schwartz 1992; Smith and Schwartz 1997). Some of the shortcomings with the published methods are minor, for example, conducting a components analysis rather than a factor analysis as if the data contained no measurement error (Hofstede, Bond, and Luk 1993) or conducting an orthogonal rather than oblique rotation as if factors were not expected to be correlated (Hofstede 1983). Perhaps more problematic is that only descriptive techniques are employed. In our modeling approach, hypothesis tests are available for stronger conclusions.

An especially peculiar approach prevalent in the cross-cultural literature is to conduct a so-called "ecological"

factor analysis (Hofstede 1983) wherein a matrix of means, with countries as rows and measured variables as columns, is factor analyzed (Bond 2002).[2] Once the matrix of means has been computed, the countries constitute the observations in the processed data to be factored, and the sampling of countries may influence and limit the results. If, say, only Asian countries are sampled, correlations may be weak across certain variables, due to range restrictions on variances. If Asian and Latin American countries are sampled, correlations between some variables may appear spuriously strong, due to the lack of European country inclusion. These problems are manifest in the overrepresentation and underrepresentation of cultures. In our modeling, the sampling of both individual customers and countries or cultures can remain intact.

THE DATA

In this specific research project, we have data from 7,378 respondents who are former or current customers from a database maintained based on a business-to-business transaction with a U.S.–based global multinational firm. These business customers are likely to be somewhat more sophisticated and experienced than end-user consumers. These respondents were drawn from a sampling scheme of database marketing relationships between a high-technology goods manufacturer and its revolving current customers and that for a provider of a financial service.

The data represent 21 countries in the evaluation of the purchase of the high-tech good, with a total sample size of 2,063 and an average sample size per country of 98.2. A subset of 12 of those countries is represented in the services marketplace, with a total sample size of 5,315 and an average sample size of 442.9. Thus, the data demonstrate different strengths: More countries are represented for goods, but there are larger sample sizes per country per services. Representatives of an objective customer satisfaction marketing research firm conducted the survey interviews in these countries, for which surveys were translated and back translated to ensure data quality. Local survey experts were sent to contact the past and current customers from the sampling frame of the database from the focal global firm. Data were collected country by country, but our analyses aggregated the data to the more abstract level of the theoretical culture.

Theoretical parsimony encourages the aggregation across countries within a similar overarching culture. Our sets of cultural states—namely, Latin America (Argentina, Brazil, Colombia, Mexico, Peru, Puerto Rico), Northern Europe (France, Germany, Netherlands, Switzerland), Southern Europe (Italy, Spain), and Asia (Hong Kong, Singapore, Taiwan) are consistent with the findings of other researchers who have explored multiple countries simultaneously (cf. Brake, Walker, and Walker 1995; Hofsted 1983; Schwartz 1992; Smith, Dugan, and Trompenaars 1996).[3]

Hence, in this article, we examine customer evaluation judgments for two U.S.–based global multinational firms with presences in 21 countries, one a goods manufacturer, another a service provider. Our data comprise a two-way factorial structure: cultures (i.e., Latin America, Northern and Southern Europe, and Asia) crossed with marketplace offerings (goods vs. services). Accordingly, we will entertain hypotheses about whether customer evaluations, for example, judgments of the likelihood of repeat purchasing, are a function of culture differences, sector differences, or both (e.g., is there an interaction where value determines repeat purchase in Asia for goods but quality does for services, whereas another pattern characterizes decisions of customers in another culture?).

SEM IN A FACTORIAL DESIGN

In this section, we present the techniques for testing SEM across cultures, industries, and the interactions between the two. To begin, we briefly review multigroup factor analysis and SEM.

There exists a stream of research on "multigroup factor analysis" (Jöreskog 1971; Tucker 1974) and tests of "factor invariance" across populations (Buss 1975; Cattell 1978; Cudeck and Browne 1983; Reise, Widaman and Pugh 1993; Tanaka and Bentler 1983), which address a cross-validation-like question, "To what extent does the factor pattern observed in one population resemble that found in another?" We will review it briefly to familiarize the reader with the modeling approach and logic. However, that literature addresses the extension neither to the structural parameters nor to factorial data sets, both of which we then present.

A Review of Multigroup Factor Analysis

As per the classic factor analysis model, we denote $\mathbf{x}$ as the p-variate vector per observation; μ is also $p \times 1$ (though often the data are mean-centered); the covariance matrix is Σ, $p \times p$, and the data are modeled

2. Researchers note these correlations are stronger than data on individuals (Hofstede et al. 1990). They must be, given the greater reliability of a mean over a raw data point. However, this method is problematic unless the factor model is modified for the change in unit of analysis (distributional assumptions will be violated, rendering fit statistics invalid).

3. The United Kingdom will be treated as distinct from Europe (cf. Campbell et al. 1988) and Japan from Asia (Graham et al. 1988). Australia and New Zealand will be aggregated, and the United States will be distinct.

$$x = \mu + \Lambda f + z. \quad (1)$$

The standard factor analytic model of sigma is,

$$\Sigma = \Lambda\Phi\Lambda' + \Psi^2, \quad (2)$$

where Λ is the $p \times r$ matrix of factor loadings, Φ the $r \times r$ matrix of factor intercorrelations, and Ψ^2 the $p \times p$ matrix of uniqueness variances.

Jöreskog (1971) presented the case for which there exist p_g measures on N_g number of subjects in population group g, denoting degrees of freedom per group, $n_g = N_g - 1$, with a total $n = \Sigma n_g$, extracting r_g factors in group g. All parameters are allowed to vary (depend on g) in the full form of the multigroup factor model:

$$x_g = \mu_g + \Lambda_g f_g + z_g, \quad (3)$$

and,

$$\Sigma_g = \Lambda_g \Phi_g \Lambda_g' + \Psi_g^2. \quad (4)$$

The $k_g = p_g r_g + 2r_g(r_g + 1) + p_g$ estimable parameters make up $\theta_g(\Lambda_g, \Phi_g, \Psi_g)$ per group or, universally, $\theta' = (\theta_1', \theta_2', \ldots, \theta_m')$. The standard fit function is relaxed to allow for the parametric dependency upon population (g):

$$F = \tfrac{1}{2} \sum_{g=1}^{m} n_g \left[\log\left|\Sigma_g\right| + tr\left(S_g \ \Sigma_g^{-1} \right) - \log\left|S_g\right| - p_g \right]. \quad (5)$$

F is a function of the variant population fits, $G(\pi_i)$, which may be further simplified by defining $K = \{k_{ij}\}$, which is s (number of nonzero parameters) by t (additional constraints on some parameters to be equal across groups), $k_{ij} = 1$ if $\pi_i = k_i$; otherwise, $k_{ij} = 0$, so fitting is a simple summative function: $\partial H/\partial k = K'(\partial G/\partial \pi)$.

Jöreskog (1971) presented a series of hypotheses that test in an increasingly constrained manner what aspects of the factor analytical model may be constant across the populations. He began with the omnibus hypothesis of equality of covariance matrices: H_Σ: $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_m$. He suggested that if the analyst cannot reject this hypothesis, modeling should continue using an aggregate of the groups' data, and continued speculation into group differences would be pointless. Continuing, however, he offered a number of hypotheses. The least restrictive manner for factor analyses to be statistically similar would be for their ranks to be equal: H_r: $r_1 = r_2 = \ldots = r_m$. If the groups are identical in the number of factors to be extracted, the factor loadings matrices may also be tested for their comparability: H_Λ: $\Lambda_1 = \Lambda_2 = \ldots = \Lambda_m$.

If the factor loadings are similar within sampling variability, uniqueness factors may be examined: $H_{\Lambda,\Psi}$: $\Lambda_1 = \Lambda_2 = \ldots = \Lambda_m$ *and* $\Psi_1 = \Psi_2 = \ldots = \Psi_m$. Finally, if the groups still appear similar, factor intercorrelations may also be tested: $H_{\Lambda,\Psi,\Phi}$: $\Lambda_1 = \Lambda_2 = \ldots = \Lambda_m$, $\Psi_1 = \Psi_2 = \ldots = \Psi_m$, *and* $\Phi_1 = \Phi_2 = \ldots = \Phi_m$.

Jöreskog (1971) noted that this final hypothesis, $H_{\Lambda,\Psi,\Phi}$, is more restrictive in form than H_Σ (the equality of covariance matrices), and so $H_{\Lambda,\Psi,\Phi}$ may not fit even if H_Σ does. (If H_Σ holds, $H_{\Lambda,\Psi,\Phi}$ may not; if $H_{\Lambda,\Psi,\Phi}$ holds, H_Σ must. Conversely, if one rejects H_Σ, one would also reject $H_{\Lambda,\Psi,\Phi}$, but if one were to reject $H_{\Lambda,\Psi,\Phi}$, one might or might not reject H_Σ.)[4,5]

SEM

Jöreskog's (1971) approach is accessible in the Lisrel software, but it does not allow for a consideration of multiple populations that are created by an underlying structure, such as a factorial design. This limitation is interesting given that the data Jöreskog used as an illustration of his methodology were in fact four groups derived from a 2×2 factorial design: One factor was the class level of the elementary school children on whom the data were collected; the other factor was based on a median split of a human abilities individual differences variate. Jöreskog applied the methodology to the four groups, as if they were not linked in the factorial manner. Tanaka and Bentler (1983), in a similar use of multigroup factor analysis, also employed data arising from a 2×2 design, which were also subsequently treated as simply four independent groups. Although these treatments are not "wrong," neither are they optimal. Treating the data in this manner is analogous to conducting an analysis of variance ignoring the factorial design.

In addition, Jöreskog's (1971) multigroup model predated the explosion in popular use of SEM and its software, so another extension that should be developed is that to the structural parameters in such models, beyond these factor analytical measurement parameters. Recall that the model statements in structural equations models using Lisrel notation (Jöreskog and Sörbom 1996) includes the factor analytical measurement models for both the q exogenous ("predictor") variables, x and their constructs, ξ, and the p endogenous ("dependent") variables, y and their constructs, η:

4. In a sense, this conclusion contradicts his earlier suggestion that should the researcher fail to reject H_Σ, the researcher should not continue the modeling, because there would exist information in testing $H_{\Lambda,\Psi,\Phi}$ and its predecessor models.

5. Tucker (1974) offered a comparable modeling approach, based on classic principal axes factor analysis involving a group space and transformations.

$$x = \Lambda_x\xi + \delta, \text{ and } y = \Lambda_y\eta + \varepsilon, \quad (6)$$

as well as the structural parameters that link the exogenous constructs to the endogenous, $\Gamma(r_y \times r_x)$, and the endogenous constructs amongst themselves, $B(r_y \times r_y)$:

$$\eta = B\eta + \Gamma\xi + \zeta. \quad (7)$$

With the rows and columns of the observed covariance matrix partitioned by the endogenous-exogenous distinction, as follows:

$$\Sigma = \begin{matrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{matrix}$$

the full model, inclusive of measurement and structural parameters, would resemble Equation 8:

$$\sum = \begin{bmatrix} \Lambda_y (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1\prime}\Lambda_y' + \Theta_\varepsilon & \Lambda_y (I-B)^{-1}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma'(I-B)^{-1\prime}\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{bmatrix} \quad (8)$$

Jöreskog's (1971) original multigroup factor analytical model on Λ can be extended in a rather straightforward manner to the bimeasurement models on Λ_x and Λ_y, with a few choices. For example, H_r in Jöreskog is tested via a χ^2 with $\{2[(p-r)^2 - (p+r)]\}$ degrees of freedom.[6] With measurement models on both the *x* and *y* variables, there would be variants on the rank hypotheses, for example, H_{rx}: $r_{x1} = \ldots = r_{xm}$; H_{ry}: $r_{y1} = \ldots = r_{ym}$; and collectively, H_r: H_{rx}, H_{ry}. Where the exogenous variables would be treated identically as in Jöreskog's approach, the modeler must choose among treatments of the *y* variables. In SEM, the measurement model on the *y*s is, as previously stated, $y = \Lambda_y\eta + \varepsilon$, which together with the *x*s would require a χ^2 on $2[(p + q)(p + q - 1) + r_x(r_x - 1 - 2q) + 2r_y(r_y - p)]$ degrees of freedom. Including the structural implications on the endogenous variables, for example, modeling the Σ_{yy} covariance submatrix, conditions the χ^2 on $2[(p + q)(p + q - 1) + 2(r_x - q)(r_x + 1) + 2r_y(1 - p - r_x) - r_x + 2(z_B + z_\Gamma)]$ degrees of freedom (where z_B and z_Γ are the number of fixed parameters in the *B* and Γ matrices). Finally, although unusual, it could be argued defensibly to let ψ represent the factor correlations between the endogenous constructs in a manner analogous to Φ for the exogenous. This approach would test the χ^2 on $2[(p + q)(p + q - 1) + r_x(r_x - 1 - 2q) + r_y(r_y - 1 - 2p)]$ degrees of freedom.[7]

In our multigroup structural models, we extend the Jöreskog (1971) approach, both in the direction of factorial underlying designs, and also to incorporate structural parameters. In the presentation that follows, we develop the factorial model focusing on the structural parameters, Γ and *B*, given that the resulting approach and logic is readily applicable to the measurement models on Λ_x and Λ_y.

TABLE 1
Tests of $\Sigma_1 = \Sigma_2$

	χ^{2a}	*RMSEA*	*GFI*
Asia			
Σ(Hong Kong) = Σ(Taiwan)?	271.0	.035	.97
Hong Kong = Singapore?	391.0	.052	.94
Taiwan = Singapore?	390.0	.048	.93
Japan			
Japan = Hong Kong?	482.0	.058	.93
Japan = Taiwan?	721.0	.070	.88
Japan = Singapore?	617.0	.067	.91
Europe			
France = Germany?	337.0	.043	.95
Italy = Spain?	401.0	.047	.93
United Kingdom			
France = United Kingdom?	378.0	.048	.94
Italy = United Kingdom?	357.0	.045	.94
United States			
United States = Canada?	741.0	.089	.86
Australia and New Zealand = Asia?	214.0	.067	.97
Australia and New Zealand = Europe?	203.0	.040	.99
Australia and New Zealand = United States and Canada?	177.0	.038	.99
Discriminant validity tests			
Brazil = Germany?	587.0	.063	.89
Italy = Singapore?	474.0	.055	.90

NOTE: RMSEA = root mean square error of approximation; GFI = goodness-of-fit index.
a. For all χ^2s, 44 degrees of freedom. $p < .001$.

Aggregating Covariance Matrices via Testing H_Σ

Using logic analogous to Jöreskog's (1971), we first attempted to test the hypothesis of equal covariance matrices, H_Σ (stated previously), as a benchmark. However, with 12 countries represented in the services sector and 21 for the data on the high-technology good, there were constraints of insufficient memory. We began to aggregate our covariance matrices into a reduced number of common cultures to make the size of the data problem manageable by testing pairwise whether doing so would be empirically defensible.

Table 1 contains the indices of fit: chi-square, root mean square error of approximation (RMSEA), and good-

6. These *df* are computed as usual: the number of data points, $2p(p + 1)$, minus the number of parameters estimated, $pr + r(r + 1)/2 + p$ (minus constraints put on the estimates, r^2), for $2p(p + 1) - [pr + r(r + 1)/2 + p - r^2]$.

7. Derivations of these *df* are available from the first author.

ness of fit indices (GFI) for these pairwise tests of H_0: $\Sigma_1 = \Sigma_2$. The chi-squares (in the left column) are uniformly large, which would indicate the hypothesis of $\Sigma_1 = \Sigma_2$ is suspect; however, they are probably large as a function of sample sizes. At the other extreme, the GFIs (in the right column) are predominately large, which would indicate the hypothesis $\Sigma_1 = \Sigma_2$ is tenable for many of the pairs of covariance matrices. The RMSEA index (in the center column) is more diagnostic simply because it varies in the information it conveys across the tests conducted, so we will focus on it as our guiding index.

The first set of three tests indicate that the data in Hong Kong, Singapore, and Taiwan might be aggregated to form an Asian culture group with little loss of distinct information—none of the RMSEAs, when rounded, exceed .05 (e.g., $\Sigma_{HongKong}$ would predict $\Sigma_{Singapore}$ reasonably well). In contrast, the next three tests suggest that Japan should remain distinct from these other Asian countries, given that all the RMSEAs exceed .05 (i.e., the null hypothesis of $\Sigma_{Japan} = \Sigma_{Asia}$ does not fit). The next two tests suggest it will be safe to collect the Northern European countries and the Southern European countries. The two tests that follow suggest we could proceed by treating the U.K. data with Europe and perhaps the Northern and Southern European countries together as well. The United States does not appear much like Canada, and Australia and New Zealand appears closer to Europe and the United States than to their geographically proximal Asian neighbors.

Given that all the RMSEAs investigated thus far are roughly near .05, albeit on the proper side (indicating fit or lack of fit), we considered it useful to test pairs of countries that should not be aggregated on theoretical grounds. In the spirit of discriminant validity, when comparing Brazil and Germany or Italy and Singapore, the RMSEAs indicate their respective covariance matrices should not be treated as identical.

The Factorial Structural Equations Model (FAC-SEM)

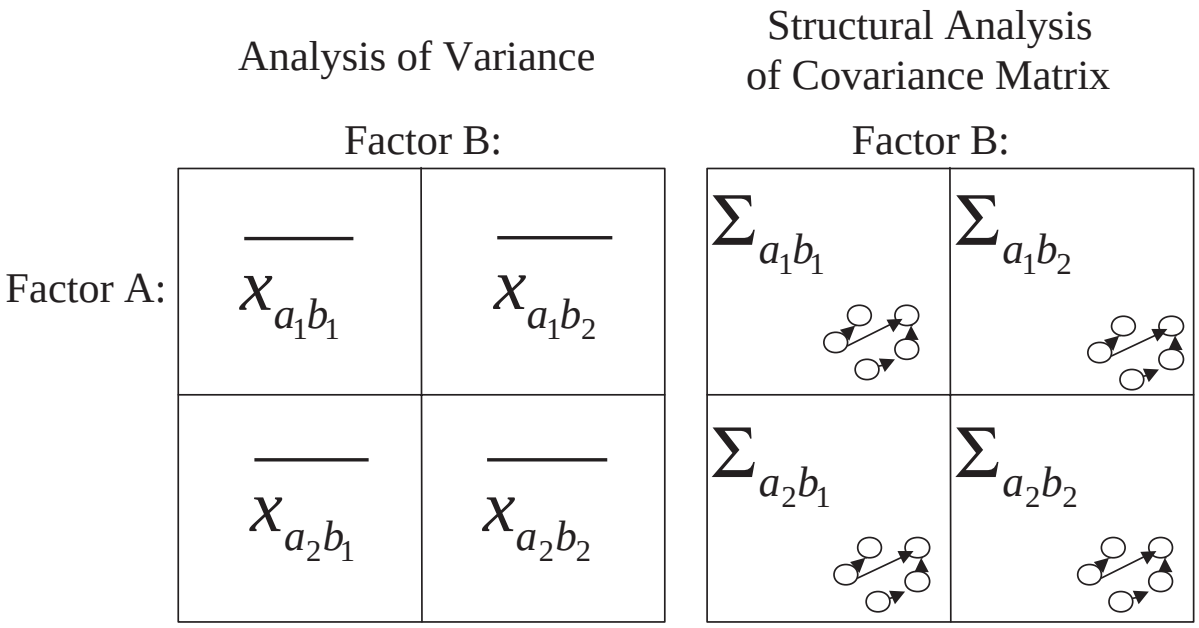
In our factorial, we will fit structural models to the covariance matrices, and our inquiry is to test whether some structural parameter(s), γ or β, vary significantly across the factors of marketplace and culture. To do this, we extend a factorial experimental design logic to SEM for covariance matrices. Our pursuit is the theoretical converse of Bagozzi and Yi (1989, 1992) or Hauser and Goldberger (1971): They fit ANOVA data via SEM, whereas we are extending SEM for factorial data using the logic of ANOVA. Specifically, they use exogenous dummy variables to represent factors, much as one would to conduct an ANOVA via a regression model; the difference between regression and their model is that theirs explicitly incorporates measurement error. Our problem is more complex, not the reworking of one model (structural equations) to try to improve on another (ANOVA). We have the SEM depicted in Figure 1, which we desire fit to a factorial structure, as depicted in Figure 2, where each condition is comprised of a combination of sector and cultural group, and instead of yielding means, as per ANOVA, we have covariance matrices to be fit and contrasted.

The goals of our model are not unrelated to the structural equations approach to multitrait multimethod matrices (Cudeck and Browne 1983), wherein methods and traits factors give rise to observed indicators. However, by definition those models do not contain interaction terms, and the modification would require the inclusion of a product term in an otherwise nonmultiplicative form, hierarchical or orthogonal factors if sufficient degrees of freedom existed, or correlated errors, a typical last-ditch and nontheoretical resolution. We sought a statistically more elegant solution.

Other models that have similar elements include Muthén (1989); Yung (1997); and Jedidi, Jagpal, and DeSarbo (1997). Muthén, like Jöreskog (1971) previously described, focused on the measurement component of an SEM, though he begins to incorporate the possibility of a factorial mixture in the sample data being modeled, for example, subgroups formed by race, gender, and so on. Yung and Jedidi, Jagpal, and DeSarbo focused on the measurement factors also but took on the challenging issue of fitting models and identifying segments within the sample simultaneously, as per the currently popular latent class modeling approaches. A future step may be to combine these classes of models to enable the fitting of covariance matrices whether the group membership is known a priori or not. Marsh (1994) offered a factorial examination of confirmatory factor analysis, and our research may said to be a further extension still to include the structural part of the typical structural equations modeling approach. We might state a "plug" though for our methods, in that our models may be fit using Lisrel software, with no extensive programming modification, whereas these other techniques require special software for fitting.

In our factorial structural equations modeling approach, we state hypotheses regarding the sufficient statistics of the covariance matrices and their derived structural (and measurement) parameters, for example, H_0: $\Sigma_i = \Sigma_{i'} = \Sigma \forall i,i'$. To keep the following presentation simple, we created scales by averaging over items indicating a common construct (as if we had a path model with no measurement elements). We also dispense with the measurement issues because, as we have stated, they have already been addressed in the literature via multigroup factor analysis models, and our contribution lies with the extension of this logic and the development of the models to situations of

FIGURE 2
The Analytical Paradigm and Data Scenario



factorial designs and the structural elements of the model. Thus, in fitting the models that follow, we posit not the typical measurement models as per Equation 6, recall:

$$x = \Lambda_x\xi + \delta \text{ and } y = \Lambda_y\eta + \varepsilon;$$

but for simplicity (which may be relaxed easily):

$$x = \xi, \text{ and } y = \eta; \tag{9}$$

that is,

$$\Lambda_x = I_q,\ \Lambda_y = I_p,\ \Theta_\delta = 0_q, \text{ and } \Theta_\varepsilon = 0_p.$$

The covariance submatrix elements of Equation 8,

$$\Sigma_{yy} = \Lambda_y(I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I - B)^{-1'}\Lambda_y' + \Theta_\varepsilon$$

$$\Sigma_{xy} = \Lambda_x\Phi\Gamma'(I - B)^{-1'}\Lambda_y'$$

$$\Sigma_{xx} = \Lambda_x\Phi\Lambda_x' + \Theta_\delta,$$

then may be rewritten to reflect the simplified measurement assumptions:

$$\Sigma_{yy} = (I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I - B)^{-1'}$$

$$\Sigma_{xy} = \Phi\Gamma'(I - B)^{-1'}$$

$$\Sigma_{xx} = \Phi. \tag{10}$$

Doing this allows us to focus on the structural parameters, B and Γ.

Each time we test a variant of B or Γ, we test three omnibus hypotheses: one for the main effect of culture, one to compare goods versus services, and one for the interaction of these factors. We proceed in a branching hierarchical fashion, first positing the most restrictive hypothesis—that the hypotheses hold for both B and Γ. We will then constrain the structural parameters separately, first testing the invariance of the Bs but allowing Γs to vary across samples, then testing the Γs without constraining the Bs.

The interaction hypothesis of invariance for both B and Γ is stated,

$$H_{interaciton,\,B,\Gamma}: \quad H_{B-culture\times goods/services}: B_{i,\,j} = B_{i',\,j'}\ \forall\, i,\, j,$$
$$\text{and}\quad H_{\Gamma-culture\times goods/services}: \Gamma_{ij} = \Gamma_{i'j'}\ \forall\, i,\, j,$$

where B_{ij} is the matrix of structural parameters estimated to represent effects among the endogenous factors, Γ_{ij} is that which connects the exogenous to the endogenous, each in the (ith, jth) sample, where $i = 1, 2$ representing the sectors goods or services, $j = 1, 2, 3, 4$ representing the cultures Latin America, Northern Europe, Southern Europe,

and Asia. If this null held, it would reflect itself in the structural variant of Equation 7,

$$\eta_{ij} = B\eta + \Gamma\xi + \zeta_{ij},$$

or, in terms of the covariance submatrices (note the lack of subscripts, or dependence upon, *i* and *j*):

$$\Sigma_{yy}^{(ij)} = (I - B)^{-1}(\Gamma\Phi_{ij}\Gamma' + \Psi_{ij})(I - B)^{-1'}$$

$$\Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma'(I - B)^{-1'}$$

$$\Sigma_{xx}^{(ij)} = \Phi_{ij}. \qquad (11)$$

The main effect hypothesis for the comparison of goods to services, with invariance for both *B* and Γ, is stated,

$$H_{goods/services,\,B,\Gamma}: \qquad H_{B-goods/services}: \; B_{1.} = B_{2.} = B_{i.},$$
$$\text{and} \quad H_{\Gamma-goods/services}: \qquad \Gamma_{1.} = \Gamma_{2.} = \Gamma_{i.},$$

where a period in the subscript represents the aggregating operator (here over cultures, B_{ij} is assumed to be homogeneous over *j*, for each *i*, and then the statistic tests whether $B_{1.} = B_{2.}$). If this hypothesis held, the model would reflect,

$$\eta_{ij} = B_{i.}\eta + \Gamma_{i.}\xi + \zeta_{ij},$$

and, in terms of the covariance submatrices,

$$\Sigma_{yy}^{(ij)} = (I - B_{i.})^{-1}(\Gamma_{i.}\Phi_{ij}\Gamma_{i.}' + \Psi_{ij})(I - B_{i.})^{-1'}, \qquad (12)$$
$$\Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma_{i.}'(I - B_{i.})^{-1'}, \text{ and } \Sigma_{xx}^{(ij)} = \Phi_{ij}.$$

Analogously, the main effect hypothesis for cultures, assuming invariance for both *B* and Γ, is,

$$H_{cultures,\,B,\Gamma}: \; H_{B-culture}: \; B_{.1} = B_{.2} = B_{.3} = B_{.4} = B_{.j},$$
$$\text{and} \quad H_{\Gamma-cultures}: \; \Gamma_{.1} = \Gamma_{.2} = \Gamma_{.3} = \Gamma_{.4} = \Gamma_{.j},$$

with the model,

$$\eta_{ij} = B_{.j}\eta + \Gamma_{.j}\xi + \zeta_{ij}$$

and the covariance matrices,

$$\Sigma_{yy}^{(ij)} = (I - B_{.j})^{-1}(\Gamma_{.j}\Phi_{ij}\Gamma_{.j}' + \Psi_{ij})(I - B_{.j})^{-1'}, \qquad (13)$$
$$\Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma_{.j}'(I - B_{.j})^{-1'}, \text{ and } \Sigma_{xx}^{(ij)} = \Phi_{ij}.$$

With these models, the interactions and main effects constraining both *B* and Γ are fit, so hierarchically we proceed to constrain the structural parameters separately, first testing hypotheses on *B*, but allowing Γs to vary across samples:

Variants of H_B:

$$H_{B-goods/services}: \qquad B_{1.} = B_{2.} = B_{i.},$$
$$H_{B-culture}: \qquad B_{.1} = B_{.2} = B_{.3} = B_{.4} = B_{.j},$$
$$H_{B-culture\times goods/services}: \; B_{i,j} = B_{i',j'} \; \forall\, i, j.$$

The hypothesis $H_{B\text{goods/services}}$ posits the model and covariance matrices (note the Γ_{ij}):

$$\eta_{ij} = B_{i.}\eta + \Gamma_{ij}\xi + \zeta_{ij},$$

$$\Sigma_{yy}^{(ij)} = (I - B_{i.})^{-1}(\Gamma_{ij}\Phi_{ij}\Gamma_{ij}' + \Psi_{ij})(I - B_{i.})^{-1'},$$

$$\text{and } \Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma_{ij}'(I - B_{i.})^{-1'}. \qquad (14)$$

The hypothesis $H_{B\text{culture}}$ posits,

$$\eta_{ij} = B_{.j}\eta + \Gamma_{ij}\xi + \zeta_{ij},$$

$$\Sigma_{yy}^{(ij)} = (I - B_{.j})^{-1}(\Gamma_{ij}\Phi_{ij}\Gamma_{ij}' + \Psi_{ij})(I - B_{.j})^{-1'},$$

$$\text{and } \Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma_{ij}'(I - B_{.j})^{-1}. \qquad (15)$$

The interaction hypothesis $H_{B\text{culture}\times\text{goods/services}}$ posits,

$$\eta_{ij} = B\eta + \Gamma_{ij}\xi + \zeta_{ij},$$

$$\Sigma_{yy}^{(ij)} = (I - B)^{-1}(\Gamma_{ij}\Phi_{ij}\Gamma_{ij}' + \Psi_{ij})(I - B)^{-1'},$$

$$\Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma_{ij}'(I - B)^{-1'}. \qquad (16)$$

Conversely, we then fit H_Γ without constraining the *B*s:

Variants of H_Γ:

$$H_{\Gamma-goods/services}: \qquad \Gamma_{1.} = \Gamma_{2.} = \Gamma_{i.},$$
$$H_{\Gamma-culture}: \qquad \Gamma_{.1} = \Gamma_{.2} = \Gamma_{.3} = \Gamma_{.4} = \Gamma_{.j},$$
$$H_{\Gamma-culture\times goods/services}: \; \Gamma_{i,j} = \Gamma_{i'j'} \; \forall\, i, j.$$

For $H_{\Gamma\text{-goods/services}}$, we fit (note the B_{ij}),

$$\eta_{ij} = B_{ij}\eta + \Gamma_{i.}\xi + \zeta_{ij},$$

$$\Sigma_{yy}^{(ij)} = (I - B_{ij})^{-1}(\Gamma_{i.}\Phi_{ij}\Gamma_{i.}' + \Psi_{ij})(I - B_{ij})^{-1'},$$

$$\text{and } \Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma_{i.}'(I - B_{ij})^{-1'}. \qquad (17)$$

For $H_{\Gamma culture}$, we fit,

$$\eta_{ij} = B_{ij}\eta + \Gamma_{.j}\xi + \zeta_{ij},$$

$$\Sigma_{yy}^{(ij)} = (I - B_{ij})^{-1}(\Gamma_{.j}\Phi_{ij}\Gamma_{.j}' + \Psi_{ij})(I - B_{ij})^{-1'},$$

$$\text{and } \Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma_{.j}'(I - B_{ij})^{-1'}. \quad (18)$$

And finally, for $H_{\Gamma culture \times goods/services}$, we fit,

$$\eta_{ij} = B_{ij}\eta + \Gamma\xi + \zeta_{ij},$$

$$\Sigma_{yy}^{(ij)} = (I - B_{ij})^{-1}(\Gamma\Phi_{ij}\Gamma' + \Psi_{ij})(I - B_{ij})^{-1'},$$

$$\text{and } \Sigma_{xy}^{(ij)} = \Phi_{ij}\Gamma'(I - B_{ij})^{-1'}. \quad (19)$$

Finally, simply to demonstrate additional details of this approach, we fit a univariate test on a particular beta, β_{12}:

$$H_{\beta-goods/services}: \quad (\beta_{12})_{1.} = (\beta_{12})_{2.}, \quad (20)$$
$$H_{\beta-culture}: \quad (\beta_{12})_{.1} = (\beta_{12})_{.2} = (\beta_{12})_{.3} = (\beta_{12})_{.4},$$
$$H_{\beta-culture \times goods/services}: (\beta_{12})_{ij} = (\beta_{12})_{i'j'}\ \forall i, j.$$

Operationally, we fit the model depicted in Figure 1 to the covariance matrices, Σ_{ij}s, applying the cross-validation (multisample) function programmed, for example, in Lisrel. To test $H_{B,\Gamma}$ in the interactions, we constrained B and Γ to be invariant across all eight covariance matrices—all four cultures, both sectors. To test for the main effect of culture, we allowed there to be estimated four $B_{.i}$s and/or $\Gamma_{.i}$s (depending on the variation of the hypothesis being tested), one for each culture, with there being assumed homogeneity across the sectors; for example, the B and Γ for Asian ratings of goods was equated with those for the Asian ratings of services but were allowed to vary from the Latin ratings of goods (and services), and so on, and then the invariance of the four $B_{.i}$s (or $\Gamma_{.i}$s) was tested. Conversely, to test for the main effect of sector, we allowed the estimation of a "goods" B and Γ and a "services" B and Γ but constrained equality across cultures; for example, $B_{\text{LatinAmerica}} = B_{\text{NEurope}} = B_{\text{SEurope}} = B_{\text{Asia}}$ for goods, and $B^*_{\text{LatinAmerica}} = B^*_{\text{NEurope}} = B^*_{\text{SEurope}} = B^*_{\text{Asia}}$ for services, and tested whether B equaled B^*.

RESULTS

The results on these tests are presented in Table 2, using the hierarchical branching of these comparative tests. Let us first consider the main effect of marketplace and the question of whether the structural precursors for repeat purchasing of goods versus services are identical. Fitting $H_{B,\Gamma}$, we found $\chi^2_{174} = 2{,}973.99$ (all p values in Table 2 are $<.0001$). Testing H_B, we find $\chi^2_{108} = 2{,}818.15$, for a $\Delta\chi^2_{66} = 155.84$. Testing H_Γ, we find $\chi^2_{138} = 2{,}845.84$ for a $\Delta\chi^2_{36} = 128.15$. On the further exploration of $H_{\beta 12}$, we find $\chi^2_{78} = 2{,}728.48$ for a $\Delta X^2_{30} = 89.67$ (nested in the H_B model).

For the main effect tests on culture, we similarly find all effects are significant. We test $H_{B,\Gamma}$ and find $\chi^2_{129} = 3{,}515.36$, H_B yields $\chi^2_{96} = 2{,}926.06$ ($\Delta\chi^2_{33} = 589.30$). Testing H_Γ yields $\chi^2_{116} = 3{,}360.76$ ($\Delta\chi^2_{13} = 154.60$), and testing $H_{\beta 12}$ yields $\chi^2_{76} = 2{,}774.22$ ($\Delta\chi^2_{76} = 151.84$).

Similarly, all effects are significant when testing these hypotheses for the Culture × Sector interaction. $H_{B,\Gamma}$ yields $\chi^2_{191} = 3{,}975.27$, H_B, $\chi^2_{114} = 2{,}988.16$ ($\Delta\chi^2_{77} = 987.11$), testing H_Γ yields $\chi^2_{149} = 3{,}497.90$ ($\Delta\chi^2_{42} = 477.37$), and the test on $H_{\beta 12}$ results in $\chi^2_{79} = 2{,}793.33$ ($\Delta\chi^2_{35} = 194.83$).

What do all these significant tests mean? For our data, there are very clear statistical differences between the variants of the model depicted in Figure 1 required to represent the four cultures and the two sectors and furthermore their interaction; that is, understanding Asian customer evaluations of goods does not allow us to predict with certainty Southern European customer evaluations of services.

Figure 3 contains the results of the structural model fit to the eight covariance matrices. This information is the analog to presenting a table or plot of means in ANOVA; we have reported the significance tests, and we know that the models differ from cell to cell in our factorial. For each cell, the paths that are significant are depicted. As the significance tests have just indicated, the optimal models vary across the marketplace and cultural factors. In addition, we should conduct additional systematic tests because, however tempting, comparisons across the diagrams (by sector, culture, or both) should not be conducted simply by eyeballing the figures, without test statistics comparing their actual sizes, any more than one would eyeball means in ANOVA without knowing whether the seeming differences were reliable. What the χ^2 tests have told us is analogous to an F test in ANOVA of an omnibus test for a main effect or interaction. When main effects or interactions are significant, we pursue them with follow-up contrasts.

This caution is true whether the sample difference is quantitative or qualitative. For an example of the quantitative differences, consider Table 3, which contains the β_{12}, that is, for the value → continue path. The path is significant for all cultures in the purchase of the high-tech good in all four cultural markets. However, although we know these coefficients are significantly different from 0, we do not know whether they are significantly different from each other. That is, without more detailed hypothesis testing, we would not know whether these estimates were roughly (within sampling) the same, or whether perhaps

TABLE 2
Fit Statistics and Comparative Tests for β and γ Estimates

For the goods versus services main effect:

$H_{B,\Gamma}$	$\chi^2_{174} = 2{,}973.99$	
$\rightarrow H_{\Gamma}$	$\chi^2_{138} = 2{,}845.84$	$(\Delta\chi^2_{36} = 128.15)$
$\rightarrow H_{B}$	$\chi^2_{108} = 2{,}818.15$	$(\Delta\chi^2_{66} = 155.84)$
$\rightarrow H_{\beta 12}$	$\chi^2_{78} = 2{,}728.48$	$(\Delta\chi^2_{30} = 89.67)$

For the cultures main effect:

$H_{B,\Gamma}$	$\chi^2_{129} = 3{,}515.36$	
$\rightarrow H_{\Gamma}$	$\chi^2_{116} = 3{,}360.76$	$(\Delta\chi^2_{13} = 154.60)$
$\rightarrow H_{B}$	$\chi^2_{96} = 2{,}926.06$	$(\Delta X^2_{33} = 589.30)$
$\rightarrow H_{\beta 12}$	$\chi^2_{76} = 2{,}774.22$	$(\Delta\chi^2_{20} = 151.84)$

For the goods versus services by cultures interaction:

$H_{B,\Gamma}$	$\chi^2_{191} = 3{,}975.27$	
$\rightarrow H_{\Gamma}$	$\chi^2_{149} = 3{,}497.90$	$(\Delta\chi^2_{42} = 477.37)$
$\rightarrow H_{B}$	$\chi^2_{114} = 2{,}988.16$	$(\Delta\chi^2_{77} = 987.11)$
$\rightarrow H_{\beta 12}$	$\chi^2_{79} = 2{,}793.33$	$(\Delta\chi^2_{35} = 194.83)$

TABLE 3
Factorial Representation of $\hat{\beta}_{12}$, the Value → Continue Parameter Estimates (and Standard Errors)

Culture	*Market Goods*	*Market Services*	*Aggregate*
Latin America	.26 (.05)	.12 (.03)	.15 (.02)
Northern Europe	.33 (.04)	.04 (.03)	.12 (.02)
Southern Europe	.21 (.05)	.02 (.03)	.06 (.02)
Asia	.34 (.08)	–.02 (.03)	.01 (.03)
Aggregate	.28 (.03)	.04 (.01)	.07 (.01)

the Latin American .26 and the Southern European .21 are similar, yet different from the Northern European .33 and Asian .34.

Qualitative differences are actually no more easily detected. It might seem, for example, that for the purchase of the high-tech good, the presence of the significant service → continue link for the Latin American and Asian customers but not for the Northern or Southern European customers indicates they perceive the purchase differently. The estimates actually appear to be consistent with that observation: .14 (Latin), –.06 (Northern), .05 (Southern), .29 (Asian), but that conclusion is best drawn after the supportive statistics are obtained. For example, the estimate .14 is significantly different from 0, and the .05 is not, but they may not be significantly different from each other. Hence, just as we would conduct contrasts and simple effects among means in ANOVA, we conduct contrasts among the parameter estimates in FAC-SEM.

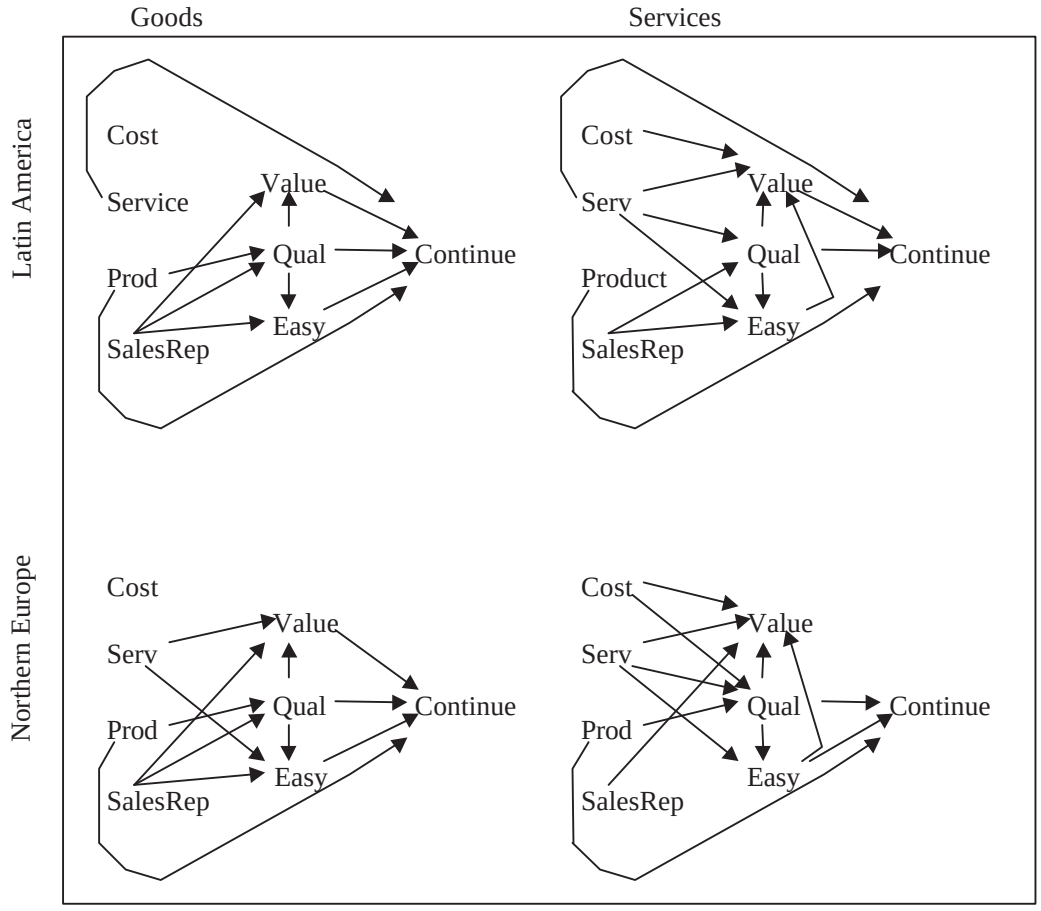
TYING THE RESULTS TO THE LITERATURE

This new statistical approach allows us to make substantive comparisons. Consider our findings in light of the marketing literature. We had predicted earlier that the cost → quality signal should be stronger in the services marketplace than for the goods (cf. Voss, Parasuraman, and Grewal 1998). In Figure 3, we see that cost tends to have significant impacts on quality and value for services, and is less likely so for goods. The services parameter (fit to the marginal covariance matrix aggregating over culture) was indeed much stronger, .093 versus .015 for goods, $t = 7.75$. In our data, the determination of the service pricing is somewhat more ambiguous (e.g., entry or user fees) and the pricing of the goods is rather competitive (so with less variance in the models, it has less impact).

Studies of customer evaluations have also demonstrated that customization can be more important than reliability in determining subsequent satisfaction and purchase behavior (Anderson, Fornell, and Lehmann 1994; Mattila 1999). Although we do not have indicator variables on customization, we might take the customer service construct as our closest proxy; and although we do not have reliability per se, we have an indicator for quality, which is usually argued to be chiefly a function of reliability (Parasuraman, Berry, and Zeithaml 1991). Aggregating over both marketplace and culture, given that this hypothesis is contingent on neither, these estimates are .29 for service → continue and .13 for quality → continue; $F = 16.01$, supporting this conceptual replication (albeit rough) that service (our "customization") is more important than quality (our "reliability") in determining repeat purchase intentions. (Admittedly, if our results had not replicated those of Anderson, Fornell, and Lehmann 1994, the difference could have been attributed to our different measures.)

We had also predicted that for the collectivistic cultures of Asia and Latin America, there should be stronger connections between the human elements in the purchase network—those representing service and sales representatives, on stated intentions of repeat purchasing. For the service → continue link, the estimates for Latin America and Asia were .14 and .29, compared with Northern and Southern European estimates of –.06 and .05; $F = 10.75$. We also see the importance of the collectivistic relational cultures on the salesreps → continue link, with estimates of .30 and .60 for Latin America and Asia, and .29 and .19 for Northern and Southern Europe; $F = 10.50$, though the estimates for Latin America and Northern Europe do not differ significantly.

FIGURE 3
Consumer Evaluation Relationship per Culture, per Marketplace



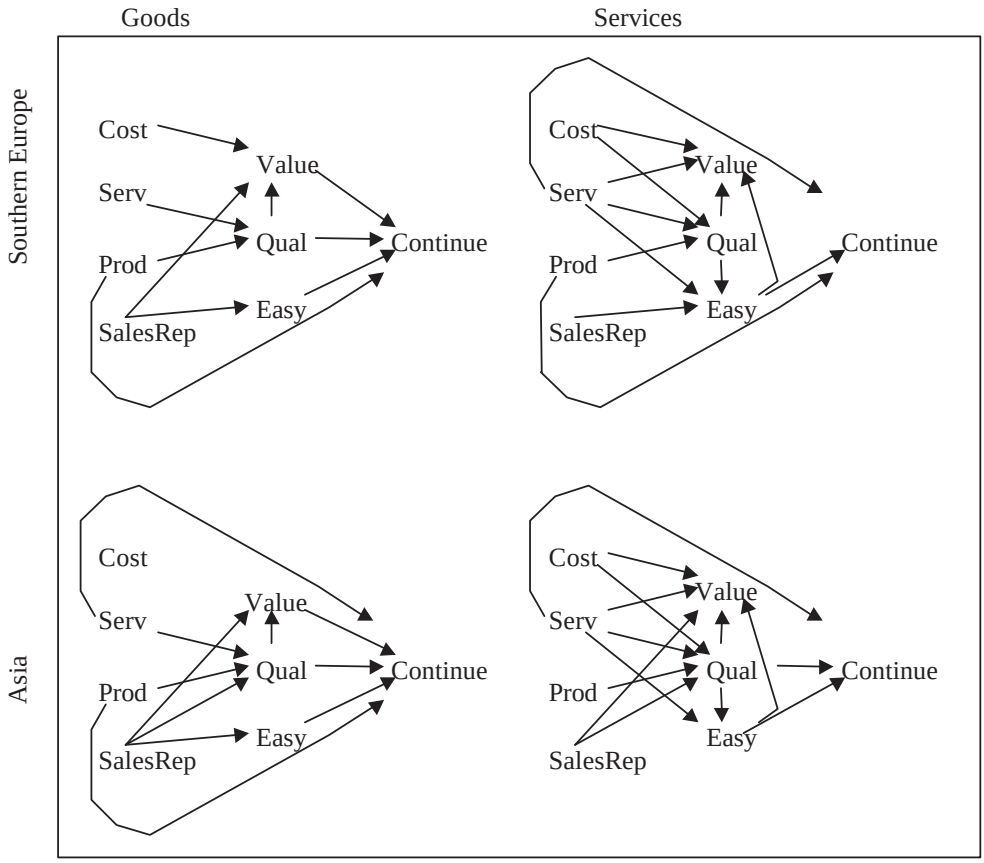
(continued)

Although these examples highlight similarities to be expected between Latin America and Asia, these two cultures should not yield identical models, given that they differ on other dimensions, for example, in that the former is thought to be more risk-averse and conservative than Asia, which is more frequently characterized as adaptive and change-seeking. This distinction might manifest itself in that the links for Latin America involving "easy to do business with" may be stronger, if one would argue that "ease" would help satisfy their conservative tendencies. This prediction holds somewhat for the purchase of the high-tech good but not for the service. For the goods, the easy → continue estimate is .23 for Latin America and .17 for Asia; $F = 4.50$. For the service, easy → continue results are –.02 and .12, significant and "in the wrong direction" (i.e., given its greater magnitude, ease of doing business is more important for the Asian). Perceptions of the easy → value path support a similar pattern, strong for Latin America for the goods (.10 vs. .02) but not for the service (.11 vs. .10).

Figure 4 introduces the comparable figures depicting significant paths for the remaining cultures that had not been included in the Latin American, European or Asia clusters: Japan, the United Kingdom, Canada, the United States, and Australia and New Zealand. We can make a few observations on each.

In the marketing literature, Song and Parry (1997) found that the successful introduction of new products in Japan was determined in large part by "product competitiveness." We can confirm that for goods, there is a significant path from product → quality, product → continue, and quality → continue. For services, only the product → quality link is supported. For Japan, the excellence of the core of the service itself is the only significant indicator of continuing to conduct business with one's partners.

FIGURE 3 (continued)

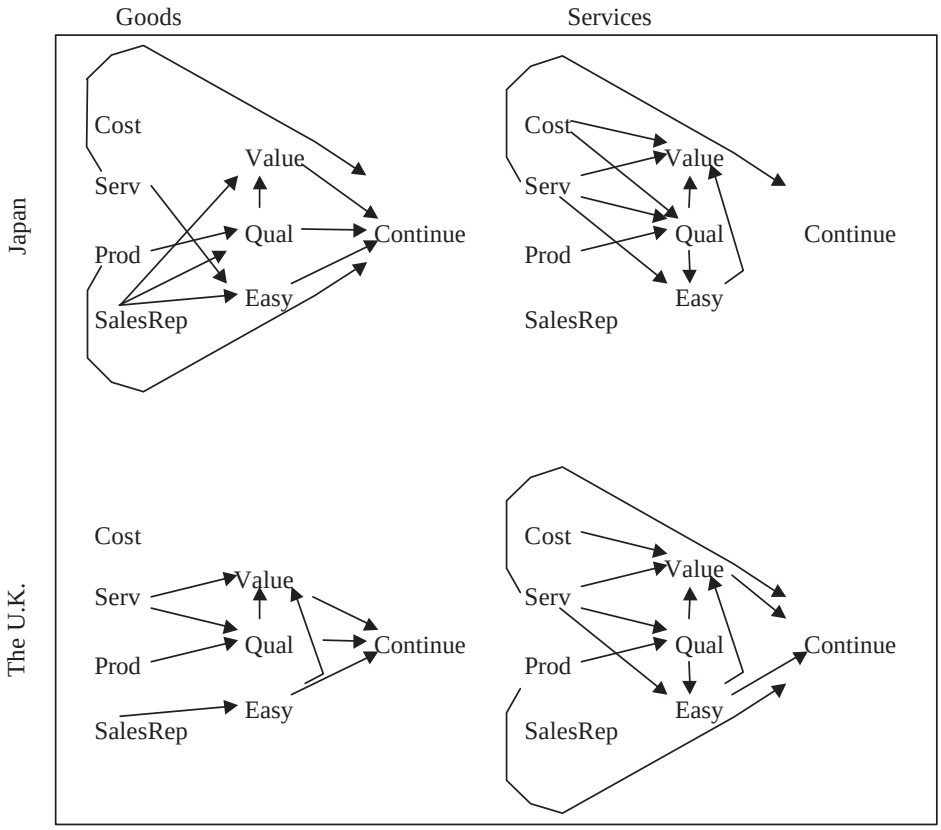


Japan also has an excellent reputation for service, with a large proportion of the worldwide top service-providing companies (Brake, Walker, and Walker 1995). One might predict that Japanese expectations would be high, with implications for the model; for example, perhaps the Japanese customer would use cost more as a cue to quality than as an indicator of value. For goods, the paths differ from each other (cost → quality is .07 and cost → value is –.10; $F = 3.43$), but neither path differs significantly from 0, so the former test, conditioning on insignificant estimates, is statistically rather unstable. For the purchase of the service, however, both paths are significantly different from 0 (cost → quality is .13 and cost → value is –.14) and they differ significantly from each other; $F = 6.28$, but both are of comparable magnitude, suggesting cost is as formidable a cue to quality as a component of value. The Japanese customer perception also varies with market; for goods, sales representatives are critical to perceptions of quality (.23), for the service, the service itself is of utmost importance (.21).

In their study on the use of foreign currency in exporting, Samiee and Anckar (1998) concluded that foreign currency choice was made to facilitate future business. We can test for a conceptual replication using our "easy to do business with" construct in place of the specific tactic of currency. We can see in Figure 3 that the easy → continue link is significant for all four cultures when purchasing the hi-tech good and for three of the cultures for purchasing the service (all but Latin America). Figure 4 confirms that the easy → continue link is nearly universally important in the other marketplaces as well (for the United Kingdom, Canada, Australia and New Zealand, the United States, and Japan, save for the service).

Regarding the other markets new to Figure 4, the nomological networks for both Canada and the United Kingdom are spartan compared with their services networks or the networks on the other cultures, evidence that we have not captured the complexities of these sets of consumers in the goods marketplace—evidently there are other drivers to which we did not have access. Data on the

FIGURE 4
Consumer Evaluation Relationship for Additional Countries, per Marketplace



(continued)

United States and Australia and New Zealand were available only in the goods market, not for the services. The United States appears to resemble Japan, Latin America, Europe, and Asia more than its English-heritages counterparts, the United Kingdom or Canada, perhaps because of the more dynamic nature of the former markets compared with the latter. Australia and New Zealand appears structurally similar to Asia and Southern Europe.

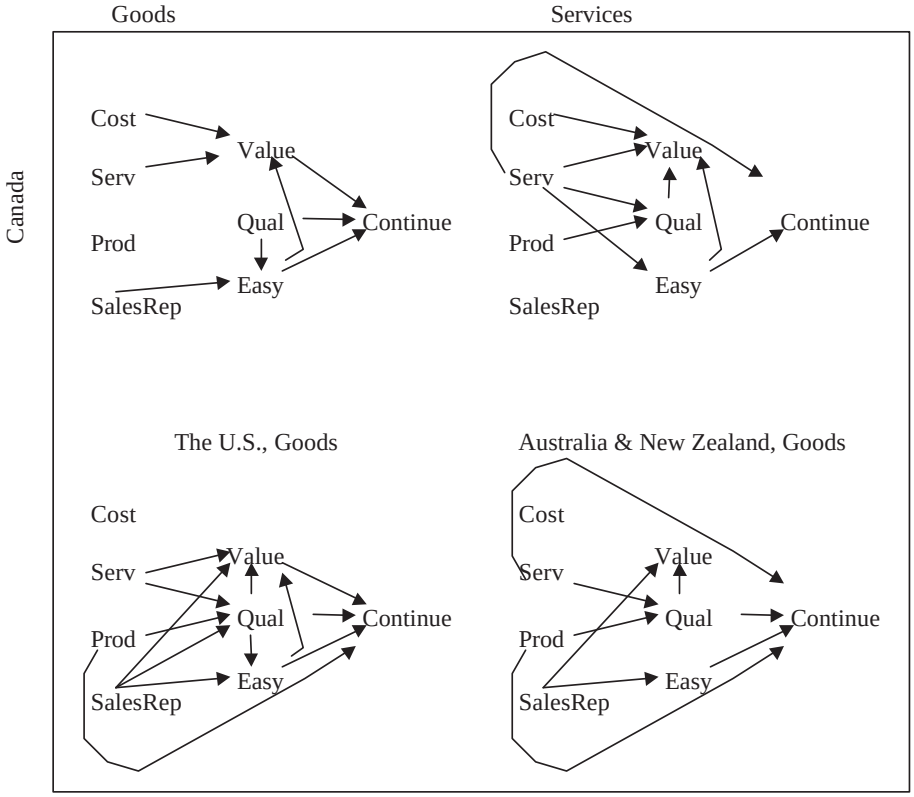
CONCLUSION

The primary intended contribution of this article is methodological in nature. We have extended the applicability of a multigroup testing logic in the context of structural equations modeling to its structural parameters and to a factorial design. The tests we created use the multigroup logic, and the follow-up contrasts are easy to compute as simple functions of elements of the output of Lisrel, Eqs, or comparable algorithms (e.g., parameter estimates, standard errors, sample sizes).

There are other applications of such a methodology. Analysis of variance would be used to compare group means, and the techniques presented in this article offer the capability of examining interrelationships between constructs. For example, a perennial issue involves the comparison between student samples and samples of "real people." Mean differences may be theoretically irrelevant if the nomological associations among constructs are stable across the groups.

One need not have groups that represent different cultures to employ these techniques. In any factorial resulting from a laboratory or field experiment, typical hypotheses involve mean differences, which are easily tested using analysis of variance. Our techniques empower the researcher to create and test hypotheses involving relationships between concepts. Any time multiple measures are obtained in an experimental setting (representing some

FIGURE 4 (continued)



network of connections, like those in Figure 1), the researcher can use these factorial multigroup extension of structural equations models, FAC-SEM, to see whether the relationships vary from group to group (as per Figures 3 and 4).

Aside from the methodological angle, we pursued theoretical and substantive investigations into culture and sector differences affecting customer evaluations. We have strengthened our understanding about cultural differences beyond simplistic main effect characterizations about cultural groups. Richer, contingent statements allow greater insight: Under certain conditions, for example, for purchases in certain product categories, certain cultural groups may behave in certain ways. Our research has begun to explore modifiers and boundary conditions on cultural statements.

We have also begun to substantiate an understanding of how customer judgments might be formed, and vary, when evaluating goods versus services. At least for these particular goods and services, sales representatives were more important for goods, and the service impacted judgments on the mediators as well as the intentions to repeat the service. The ease with which a purchase is transacted helped enhance repeat, for goods and services, but also enhanced perceptions of value for the service, which is sensible, given that services are often more effortful for the customer given the coproduction nature of a service.

These results have clear managerial implications. The marketing manager need not reinvent the wheel with the entrance into every marketplace. Offerings need to be tailored on a culture-to-culture basis, but not on the more effortful, microscopic country-to-country basis. Furthermore, given that any firm tends to specialize on offering goods or services, there can be additional simplifications; for example, the Latin American and Asian customers evaluated the high tech good more similarly than these cultures evaluated the service, so the global marketing manager subsequently introducing new technology may be safer to generalize from Asia to Latin America than the marketing manager thinking to offer a service. Or for example, sales representatives were relatively unimportant in either the goods or service offerings in the United King-

dom or Canada, presumably due in part to distribution differences. The Japanese data suggested independence of sales reps for the financial service instrument; but for the purchase of the high-tech good, the Japanese count on their sales force, more like Europe, Latin America, or Asia more broadly.

Cross-cultural inquiries and data modeling can be a challenge. We make statements in our research and in our classrooms regarding cultural differences, but often the purported differences are not statistically defended. We have offered an approach that extends the rigorous, well-known, and accessible structural equations modeling. The approach generalizes related models in the literature and is easily fit using extant software (e.g., Lisrel). The integration of the substantive questions and the means by which they might be methodologically addressed should prove useful to researchers.

REFERENCES

Anderson, Eugene W., Claes Fornell, and Donald R. Lehmann (1994), "Customer Satisfaction, Market Share, and Profitability: Findings from Sweden," *Journal of Marketing*, 58 (July), 53-66.

Bagozzi, Richard P. and Youjae Yi (1989), "On the Use of Structural Equation Models in Experimental Designs," *Journal of Marketing Research*, 26 (August), 271-84.

——— and ——— (1992), "Testing Hypotheses about Methods, Traits, and Communalities in the Direct Product Model," *Applied Psychological Measurement*, 16 (December), 373-80.

Baumgartner, Hans and Jan-Benedict E. M. Steenkamp (2001), "Response Styles in Marketing Research: A Cross-National Investigation," *Journal of Marketing Research*, 38 (May), 143-56.

Berry, John W., Ype H. Poortinga, Marshall H. Segall, and Pierre R. Dasen (1992), *Cross-Cultural Psychology: Research and Applications*. Cambridge: Cambridge University Press.

Bond, Michael Harris (2002), "Reclaiming the Individual from Hofstede's Ecological Analysis—A 20 Year Odyssey," *Psychological Bulletin*, 128 (1), 73-77.

Boulding, William, Richard Staelin, Ajay Kalra, and Valarie A. Zeithaml (1993), "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions," *Journal of Marketing Research*, 30, 7-27.

Brake, Terence, Danielle Medina Walker, and Thomas Walker (1995), *Doing Business Internationally: The Guide to Cross-Cultural Success*. New York: McGraw Hill.

Buss, Albert (1975), "An Inferential Strategy for Determining Factor Invariance across Different Individuals and Different Variables," *Multivariate Behavioral Research*, 10 (July), 365-71.

Campbell, Nigel C. G., John L. Graham, Alain Jolibert, and Hans Gunther Meissner (1988), "Marketing Negotiations in France, Germany, the United Kingdom, and the United States," *Journal of Marketing*, 52 (April), 49-62.

Cattell, Raymond B. (1978), "Matched Determiners vs. Factor Invariance: A Reply to Korth," *Multivariate Behavioral Research*, 13 (October), 431-48.

Cooper, Catherine R. and Jill Denner (1998), "Theories Linking Culture and Psychology: Universal and Community-Specific Processes," *Annual Review of Psychology*, 49, 559-84.

Craig, C. Samuel and Susan P. Douglas (2000), *International Marketing Research*, 2d ed. Chichester, UK: Wiley.

Cudeck, Robert and Michael W. Browne (1983), "Cross-Validation of Covariance Structures," *Multivariate Behavioral Research*, 18 (April), 147-67.

De Wulf, Kristof, Gaby Odekerken-Schröder, and Dawn Iacobucci (2001), "Investments in Consumer Relationships: A Cross-Country and Cross-Industry Exploration," *Journal of Marketing*, 65 (4), 33-50.

Donthu, Naveen and Boonghee Yoo (1998), "Cultural Influences on Service Quality Expectations," *Journal of Service Research*, 1 (2), 178-86.

Furrer, Olivier, Ben Shaw-Ching Liu, and D. Sudharshan (2000), "The Relationships between Culture and Service Quality Perceptions: Basis for Cross-Cultural Market Segmentation and Resource Allocation," *Journal of Service Research*, 2 (4), 355-71.

Graham, John L., Dong Ki Kim, Chi-Yuan Lin, and Michael Robinson (1988), "Buyer-Seller Negotiations around the Pacific Rim," *Journal of Consumer Research*, 15 (June), 48-54.

Greenleaf, Eric A. (1992), "Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles," *Journal of Marketing Research*, 29 (May), 176-88.

Hauser, Robert M. and Arthur S. Goldberger (1971), "The Treatment of Unobservable Variables in Path Analysis," in *Sociological Methodology*, Herbert L. Costner, ed. San Francisco: Jossey-Bass, 81-177.

Hofstede, Geert (1983), "Dimensions of National Cultures in Fifty Countries and Three Regions," in *Expectations in Cross-Cultural Psychology*, J. B. Deregowski, S. Dziurawiec, and R. C. Annis, eds. Lisse, Switzerland: Swets & Zeitlinger, 335-55.

———, Michael Harris Bond, and Chung-leung Luk (1993), "Individual Perceptions of Organizational Cultures: A Methodological Treatise on Levels of Analysis," *Organization Studies*, 14 (4), 483-503.

———, Bram Neuijen, Denise Daval Ohayv, and Geert Sanders (1990), "Measuring Organizational Cultures: A Qualitative and Quantitative Study across Twenty Cases," *Administrative Science Quarterly*, 35, 286-316.

Jedidi, Kamel, Harsharanjeet S. Jagpal, and Wayne S. DeSarbo (1997), "Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity," *Marketing Science*, 16 (1), 39-59.

Jöreskog, Karl G. (1971), "Simultaneous Factor Analysis in Several Populations," *Psychometrika*, 36 (4), 409-26.

———, and Dag Sörbom (1996), *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.

Klein, Jill Gabrielle, Richard Ettenson, and Marlene D. Morris (1998), "The Animosity Model of Foreign Product Purchase: An Empirical Test in the People's Republic of China," *Journal of Marketing*, 62 (January), 89-100.

Kotabe, Masaaki, Arvind Sahay, and Preet S. Aulakh (1996), "Emerging Role of Technology Licensing in the Development of Global Product Strategy: Conceptual Framework and Research Propositions," *Journal of Marketing*, 60 (January), 73-88.

Liu, Ben Shaw-Ching, Olivier Furrer, and D. Sudharshan (2001), "The Relationships between Culture and Behavioral Intentions toward Services," *Journal of Service Research*, 4 (2), 118-29.

Lonner, Walter J. and John Adamopoulos (1997), "Culture as Antecedent to Behavior," in *Handbook of Cross-Cultural Psychology*, vol. 1, *Theory and Method*, 2d ed., John W. Berry, Ype H. Poortinga, and Janak Pandey, eds. Boston: Allyn & Bacon.

Lovelock, Christopher H. (1999), *Services Marketing*, 4th ed. Upper Saddle River, NJ: Prentice Hall.

Marsh, Herbert W. (1994), "Confirmatory Factor Analysis Models of Factorial Invariance: A Multifaceted Approach, *Structural Equation Modeling*, 1 (1), 5-34.

Mattila, Anna S. (1999), "The Role of Culture in the Service Evaluation Process," *Journal of Service Research*, 1 (3), 250-61.

Mullen, Michael R. (1995), "Diagnosing Measurement Equivalence in Cross-National Research," *Journal of International Business Studies*, 26 (3), 573-93.

Muthén, Bengt O. (1989), "Latent Variable Modeling in Heterogeneous Populations," *Psychometrika*, 54 (4), 557-85.

Nakata, Cheryl and K. Sivakumar (1996), "National Culture and New Product Development: An Integrative Review," *Journal of Marketing*, 60 (January), 61-72.

Oliver, Richard L. (1997), *Satisfaction: A Behavioral Perspective on the Consumer.* New York: McGraw-Hill.

Oyserman, Daphna, Heather M. Coon, and Markus Kemmelmeir (2002), "Rethinking Individualism and Collectivism: Evaluation of Theoretical Assumptions and Meta-Analyses," *Psychological Bulletin*, 128 (1), 3-72.

Parasuraman, A., Leonard Berry, and Valarie Zeithaml (1991), "Understanding Customer Expectations of Service," *Sloan Management Review*, 39 (Spring), 39-48.

Reichheld, Frederick F. (1993), "Loyalty-Based Management," *Harvard Business Review*, 71 (2), 64-73.

Reise, Steven P., Keither F. Widaman, and Robin H. Pugh (1993), "Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance," *Psychological Bulletin*, 114 (3), 552-66.

Roth, Martin S. (1995), "The Effects of Culture and Socioeconomics on the Performance of Global Brand Image Strategies," *Journal of Marketing Research*, 32 (May), 163-75.

Rust, Roland T., Anthony J. Zahorik, and Timothy L. Keiningham (1996), *Service Marketing.* New York: HarperCollins.

Samiee, Saeed and Patrik Anckar (1998), "Currency Choice in Industrial Pricing: A Cross-National Evaluation," *Journal of Marketing*, 62 (July), 112-27.

Schwartz, Shalom M. (1992), "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries," in *Advances in Experimental Social Psychology*, vol. 25, Mark P. Zanna, ed. San Diego, CA: Academic Press, 1-65.

Smith, Peter B., Shaun Dugan, and Fons Trompenaars (1996), "National Culture and the Values of Organizational Employees: A Dimensional Analysis across 43 Nations," *Journal of Cross-Cultural Psychology*, 27 (2), 231-64.

——— and Shalom H. Schwartz (1997), "Values," in *Handbook of Cross-Cultural Psychology: Social Behavior and Applications*, vol. 3, 2d. ed., John W. Berry, Marshall H. Segall, and Cigdem Kagitçibasi, eds. Needham Heights, MA: Allyn & Bacon, 77-118.

Song, X. Michael and Mark E. Parry (1997), "The Determinants of Japanese New Product Successes," *Journal of Marketing Research*, 34 (February), 64-76.

Steenkamp, Jan-Benedict E. M. and Hans Baumgartner (1998), "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research*, 25 (June), 78-90.

Tanaka, Jeffrey S. and Peter M. Bentler (1983), "Factor Invariance of Premorbid Social Competence across Multiple Populations of Schizophrenics," *Multivariate Behavioral Research*, 18 (April), 219-30.

Triandis, Harry C. (1994), "Theoretical and Methodological Approaches to the Study of Collectivism and Individualism," in *Individualism and Collectivism: Theory, Method, and Applications*, Uichol Kim, Harry C. Triandis, Çidem Kâitçibai, Sang-Chin Choi, and Gene Yoon, eds. Thousand Oaks, CA: Sage, 41-51.

Tucker, Ledyard R. (1974), "Factor Analytic Models for the Aptitude-Achievement Distinction," in *Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement*, Donald Ross Green, ed., 85-108. New York: McGraw Hill.

Usunier, Jean-Claude (1996), *Marketing across Cultures*, 2d ed. London: Prentice Hall.

Vandello, Joseph A. and Dov Cohen (1999), "Patterns of Individualism and Collectivism across the U.S.," *Journal of Personality and Social Psychology*, 77 (2), 279-92.

Voss, Glenn B., A. Parasuraman, and Dhruv Grewal (1998), "The Roles of Price, Performance, and Expectations in Determining Satisfaction in Service Exchanges," *Journal of Marketing*, 62 (October), 46-61.

Yung, Yiu-Fai (1997), "Finite Mixtures in Confirmatory Factor Analysis Models," *Psychometrika*, 62 (3), 297-330.

Zeithaml, Valarie A. (1988), "Consumer Perceptions of Price, Quality and Value: A Means-End Model and Synthesis of Evidence," *Journal of Marketing*, 52, 2-22.

Dawn Iacobucci is a professor of marketing at the Kellogg School of Management, Northwestern University. Her research streams focus on the modeling of dyadic interactions and social networks, the conceptualization and measurement of customer satisfaction and service quality, and multivariate and methodological research questions. She is currently editor of the *Journal of Consumer Research* and recent editor of the *Journal of Consumer Psychology.* She edited "Kellogg on Marketing," "Kellogg on Integrated Marketing," "Networks in Marketing," and "Handbook of Services Marketing and Management"; and she is coauthor on Gilbert Churchill's lead text on "Marketing Research."

Douglas Grisaffe, Ph.D., is vice president and chief research methodologist for Walker Information. His responsibilities include development application and enhancement of stakeholder research products and methodologies, training in research methods, and consultation with domestic and international researchers and clients on stakeholder measurement. Doug is a professional member of the American Marketing Association.

Adam Duhachek is a doctoral student in marketing at the Kellogg School of Management, Northwestern University. His research interests include consumer coping, emotions, and methodological issues.

Alberto Marcati is a professor of management and marketing in the School of Economics and Business at the Luiss Guido Carli University in Rome, Italy. His research interests are in the areas of relationship marketing, interorganizational cooperation, and the management of innovations.