Organizational Research Methods

Introducing Functional Data Analysis to Managerial Science

Mayukh Dass and Christine Shropshire

The online version of this article can be found at:

Published by:

$SAGE

On behalf of:

RMD

The Research Methods Division of The Academy of Management

Additional services and information for *Organizational Research Methods* can be found at:

Email Alerts: http://orm.sagepub.com/cgi/alerts

Subscriptions: http://orm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://orm.sagepub.com/content/15/4/693.refs.html

>> Version of Record - Sep 21, 2012

What is This?

Introducing Functional Data Analysis to Managerial Science

Mayukh Dass[1] and Christine Shropshire[2]

Abstract

In this article, we introduce functional data analysis (FDA), a set of statistical tools developed to study information on curves or functions. We review fundamentals of the methodology along with previous applications in other business disciplines to highlight the potential of FDA to managerial science. We provide details of the three most commonly used FDA techniques, including functional principal component analysis, functional regression, and functional clustering, and demonstrate each by investigating measures of firm financial performance from a panel data set of the 1,000 largest U.S. firms by revenues from 1992 to 2008. We compare results obtained from FDA with hierarchical linear modeling and conclude by outlining ideas for future micro- and macro-level organizational research incorporating this methodology.

With the recent interest in modeling longitudinal, multigroup, and multilevel data (e.g., Bliese & Ployhart, 2002; Bou & Satorra, 2010; Misangyi, Elms, Greckhamer, & Lepine, 2006), scholars are increasingly able to test theoretical models of change in organizational phenomena that were not possible earlier. Although these methods (e.g., hierarchical linear modeling [HLM], growth models, traditional time-series models) are useful in exploring dynamics in general, they are sometimes inefficient, particularly in situations where the relationships among the variables under investigation vary over time (Ramsay & Silverman, 2005) and exhibit a complex data pattern (Jank & Shmueli, 2006). In such cases, functional data analysis, a set of statistical tools developed to study information on curves or functions, is more appropriate.

Functional data analysis (FDA) presents advantages over current techniques that make it particularly promising to organizational scholars. First, it has the ability to analyze highly nonlinear and

---

[1]Rawls College of Business, Texas Tech University, Lubbock, TX, USA
[2]Terry College of Business, University of Georgia, Athens, GA, USA

Corresponding Author:
Mayukh Dass, Rawls College of Business, Texas Tech University, Lubbock, TX 79409, USA
Email: mayukh.dass@ttu.edu

heterogeneous longitudinal data. Although HLM can handle certain nonlinear distributions, it typically restricts within-subject dynamics (i.e., Level 1 model) of all subjects to one functional structure (Singer & Willett, 2003, p. 211). In comparison, FDA allows each individual subject to determine its own functional structure, useful when dealing with highly heterogeneous data (Reddy & Dass, 2006). Second, FDA is able to handle time-varying relationships among variables and estimate their effect size and significance over time. Such analyses are not possible with HLM and growth models. Third, FDA has better forecasting power than traditional models in very dynamic environments (Dass, Jank, & Shmueli, 2011). As it incorporates the dynamic components of the environment in its functional form, FDA is able to predict the outcome with very low error relative to traditional methods. And, fourth, it is powerful in visualizing and capturing complex data patterns with a few simple measures, thus making it appropriate for the dynamism and complexity of organizational data. Given the preceding benefits, the purpose of this article is to introduce FDA to managerial science and discuss how organizational researchers can apply the method in their research.

Fundamentally, FDA and traditional models (e.g., HLM, growth models) are different as the first is performed on functional data, whereas the latter are developed on discrete data. Functional data are information that varies over a continuum of time, space, probability, or any dimension that defines data separation; examples include firm performance across time, product diversification levels or life cycle stages of firms, dynamics of organizational citizenship behavior, perceptions of organizational justice, job satisfaction, emotional affect, and so on. Longitudinal organizational data are functional, and relationships among the variables may change with time. Functional techniques allow for functional inputs and outcomes, including measures of varying intervals, at varying rates of change and in a highly nonlinear fashion, and thus can capture and measure how intra- and inter-organizational characteristics coevolve.

A functional approach enables investigations of temporal effects or the analysis of relationships of interest while addressing time dependence (e.g., Bliese & Hanges, 2004). For example, a study of new venture survival and growth following an initial public offering (IPO) could follow daily returns to assess entrepreneurial firms' "particular challenges stemming from a 'liability of newness' that heightens performance risks and makes them more strongly influenced by environmental change, competitive threats, or shifting consumer preferences" (Holcomb, Combs, Sirmon, & Sexton, 2010, p. 348). Le Mens, Hannan, and Pólos (2011) discuss complexities beyond age dependence to firm survival and argue that considerations of founding conditions and accumulation of organizational capital over time affect liabilities of newness, adolescence, and obsolescence. In line with the dynamics of organizational phenomena, FDA visually depicts and models the functional nature of data rather than singular dimensions. Studies that improve our understanding of "continuity, change and processes that unfold in various ways over time" represent high-impact contributions to organizational science because "conceptualizations and analyses of change over time were made more explicit in the models, which in turn allowed scientific examinations of the complexities involved in the change process" (Chan, 1998, p. 422). FDA includes a number of promising techniques to capture such richness and dynamics of organizational reality.

At a fundamental level, FDA is a hybrid methodology that entails both parametric and nonparametric components. It includes a nonparametric stage of discovering the mathematical form of the underlying functions and then a parametric stage to analyze them. In particular, FDA first generates a continuous and smooth curve from discrete observations, creating a functional equation to explain changes in states of a variable due to some group membership (e.g., point in time, firm or industry nesting). This flexibility makes FDA appropriate for a multitude of organizational research questions. Rather than dealing in linear space with matrix equations, FDA works with integral equations in infinite dimensions, allowing more appropriate treatment to continuous data and study of the dynamics of longitudinal changes. Research methods for macro-organizational topics have been limited to parametric or linear modeling or to those that require predetermined knowledge about

event timing, such as survival analysis or traditional spline regression. Furthermore, methods in micro-organizational research often restrict responses from study participants to one functional form and treat responses as simultaneous when they occur over an extended time period of data collection. Such restrictions may render insights from traditional methods less accurate. FDA is effectively a statistical toolkit to analyze repeated observations in exploratory, confirmatory, and predictive studies. In this article, we provide details of three commonly used FDA techniques, including functional principal component analysis, functional regression, and functional clustering. We demonstrate each technique by examining three measures of firm financial performance (return on assets [ROA], Tobin's Q, and shareholder return) from a panel data set of the 1,000 largest U.S. firms by revenues from 1992 to 2008. Moreover, we compare FDA with HLM and illustrate advantages of a functional methodology.

The rest of the article is organized as follows: We first briefly review the application of FDA to social science, including recent advances in finance, economics, marketing, and management information systems, to introduce its relevance to organizational research questions. We then provide the methodological details of FDA, demonstrate three functional techniques using financial performance measures, and present a comparative analysis between FDA and HLM. Finally, we conclude with a discussion of applications of FDA in the managerial sciences and its potential to explore a variety of micro- and macro-organizational phenomena.

Development of FDA and Previous Applications

Across a variety of disciplines, functional data analysis is a new but increasingly popular methodology, largely due to its adaptability in the multivariate context (Ramsay & Silverman, 2005, p. 399) and its ability to extend into a functional space a number of customary statistical approaches, including factor analysis, inference, classification, time series, and resampling (Ferraty & Romain, 2011). Although considering data functionally adds some complexity, researchers familiar with traditional methods such as principal components and regression analyses can easily adopt FDA to more fully capture the realities of data in managerial science, especially the multilevel and dynamic nature of organizational phenomena.

Functional Data Analysis in the Business Disciplines

As the social sciences have begun acknowledging limitations of using traditional econometric models in functional settings, FDA is gaining traction in business disciplines. Traditional models face several shortcomings that functional methods address. First, traditional methods fail to provide insights that are useful for understanding the nature of functional data. For example, consider one part of our later example: how firm size affects firm performance. We can adequately estimate the effect size with a linear model and determine a relationship between firm size and firm performance. However, as both firm performance and firm size vary over time, the econometric model will only provide results at the aggregate level, without exploring the possibility of changing relationships across time. Second, traditional methods are limited in their capability to visualize the dynamics of functions. For example, if we want to explore how changes in firm size covary with changes in firm performance and visually investigate it, econometric models are unfulfilling. Because FDA handles functional (as opposed to discrete) independent and dependent variables, researchers can model relationships between change in size and change in performance and consider predictors, controls, or outcomes as a function of nested or time-varying effects. Third, traditional methods are inadequate in situations where data suffer from missing information. For example, it is possible that some firms are missing performance data in some years within the sample period. In such cases, we have to either impute the data or consider deleting those firms in our analysis. FDA can also handle

such situations by smoothing the discovered underlying function to impute the missing values. However, FDA also offers the ability, if the signal-to-noise ratio in the data set is low and the data significantly sparse, to use information from "neighboring or similar" subjects to get more stable estimates of the missing values (Brumback & Rice, 1998; James, Hastie, & Sugar, 2000; Ramsay & Silverman, 2005). Studies published recently in business disciplines overcome the previously mentioned shortcomings of traditional methods and illustrate the effectiveness of using FDA to address organizational research questions.

Finance and Economics

Given the widespread use of financial data such as measures of risk and performance in macro-organizational studies, relevant findings from finance and economics offer insight to how FDA can illuminate our own research. Ramsay (1990) discusses limitations of traditional methods to capture the nonlinear dynamics of financial data and introduces ways to visualize data using FDA. Phase plots between the first derivative (velocity) and second derivative (acceleration) of the functions helpfully depict the dynamics underlying the data, not just predicting change but also allowing analysis of second- and third-order changes and their correlations. Cai (2011) discusses the shortcomings of traditional models such as the capital asset pricing model (CAPM) in comparison to FDA, noting a wealth of empirical evidence yet "no theoretical guidance on how betas and risk premium vary with time, or variables that represent conditioning information" (p. 178). Research predicting abnormal returns often assumes that time and market factors are uncorrelated with the error terms (Akdeniz, Altay-Salih, & Caner, 2003). Although some advances in econometric modeling allow for time-varying betas of the overall market or of an individual stock (e.g., You & Jiang, 2007), these approaches still fail to recognize covariance between the two in order to compare risk of an individual stock or portfolio against volatility of broader market indices (Cai, 2011). Functional methods offer the ability to identify trends within panel data sets and more appropriately model autoregressive time-series data, whether financial performance or other organizational phenomena.

Questions in macro-organizational research often include multilevel considerations of macroeconomic country or industry effects. Finance and economics studies point to the value of FDA in recognizing nested, dynamic organizational data. Kargin and Onatski (2008) explore the functional nature of financial data and illustrate FDA as a superior methodology due to its accurate recognition of observations as curves rather than vectors. Although they study the currency rates using daily observations over 10 years of futures contracts, their analysis reveals the broader utility and increased predictive ability of functional methods at curve forecasting, and not just in terms of market expectations for future competitive and policy conditions (see also Hong & Lee, 2003). One highly cited study using functional techniques samples the Nondurable Goods Index (NGI), a monthly indicator of production and spending patterns in the United States (Ramsay & Ramsey, 2002). Applying FDA to 100 years of data reveals reliable patterns and shifts within the NGI, and functional techniques allow visual depictions of the dynamics of variation across time through first- and second-order derivatives of the function (Ramsay & Silverman, 2002). Identification of seasonal or other cyclical trends and larger scale shifts in macroeconomic conditions, as well as how these factors change over time, has powerful implications for theory and practice for organizational scholars. Functional techniques allow researchers to address the time and maturity dimensions of data simultaneously, longitudinally, and semi- or nonparametrically.

Marketing and Management Information Systems

Findings in the marketing and management information systems disciplines also indicate how organizational researchers may use functional analyses. FDA has been shown to be a more appropriate

methodology in highly dynamic environments such as price formation in online auctions (Bapna, Jank, & Shmueli, 2008; Reddy & Dass, 2006). Bapna and colleagues (2008) find through functional data modeling that item characteristics, auction duration, and experience and ratings of the seller have varying influence on the evolution and dynamics of the price. Therefore, FDA may serve as a better alternative when testing organizational theories in dynamic environments such as the relationship between resource management and creation of value (e.g., Sirmon, Hitt, & Ireland, 2007). Similarly, marketing and management information systems studies report FDA to have stronger predictive power than traditional models in dynamically competitive environments (Dass et al., 2011; Foutz & Jank, 2010; Wang, Jank, & Shmueli, 2008), which may encourage organizational researchers to revisit and advance organizational studies in such environments (e.g., Grant, 1996).

FDA has proved useful in its ability to handle complex and large data sets, often with thousands of variables across thousands of observations (e.g., Zhu, Brown, & Morris, 2011). Yet FDA is also effective in smaller data sets, although it is important to have more subjects than covariates in the data set. Some examples include Ramsay and Silverman's (2005) analysis of the dynamics of height among 10 girls and the investigation by Reddy and Dass (2006) into price dynamics of 107 art items. Moreover, a minimum of two observations is required by FDA to generate an underlying function (Ramsay & Silverman, 2005); thus, it is flexible in terms of data set size, and its range of techniques to analyze functional data is both exploratory (e.g., functional principal components analysis) and predictive (e.g., functional regression) in nature.

Conceptual questions in organizational behavior and strategic management research regarding dynamic trends and covariance within and between observational units over time could be addressed using FDA. Whether observational units are individual employees, work teams, demographic categories, organizational divisions, companies, industries, and so on, FDA can visually depict and capture variation and patterns of longitudinal change rather than examining central tendencies. Given the evolution of functional techniques toward organizational data and the adoption of FDA by other business disciplines, we now present details of the methodology and demonstrate three FDA techniques in an exploration of measures of firm financial performance.

Method

At a fundamental level, FDA first recovers the underlying functions of the observed data and then performs different types of analyses on them. Although these functions can be subjected to any type of functional analysis, three applications of FDA have garnered particular attention: functional principal components, functional regression, and functional clustering. Functional principal components analysis (fPCA) is useful in determining the common factors or trends that are present in the dynamics of the underlying recovered functions. Functional regression examines the relationship between the recovered functions (a functional dependent variable) and predictors that affect the outcome of interest. Functional clustering establishes relationships across the functions and uses the information to group or cluster the units of analysis accordingly. A flowchart of FDA to pursue these three techniques is shown in Figure 1. Moreover, to facilitate readers applying FDA in their own research, we provide the necessary programming code in R in the appendix.

Recovering Underlying Functions

The first step of the methodology is to discover the underlying functions of the observed data. Since we may lack information regarding the nature of the underlying function of the given data, it is important that we start with a general functional form that is flexible and can be applied to most types of organizational data. A polynomial functional form is flexible and does not impose any
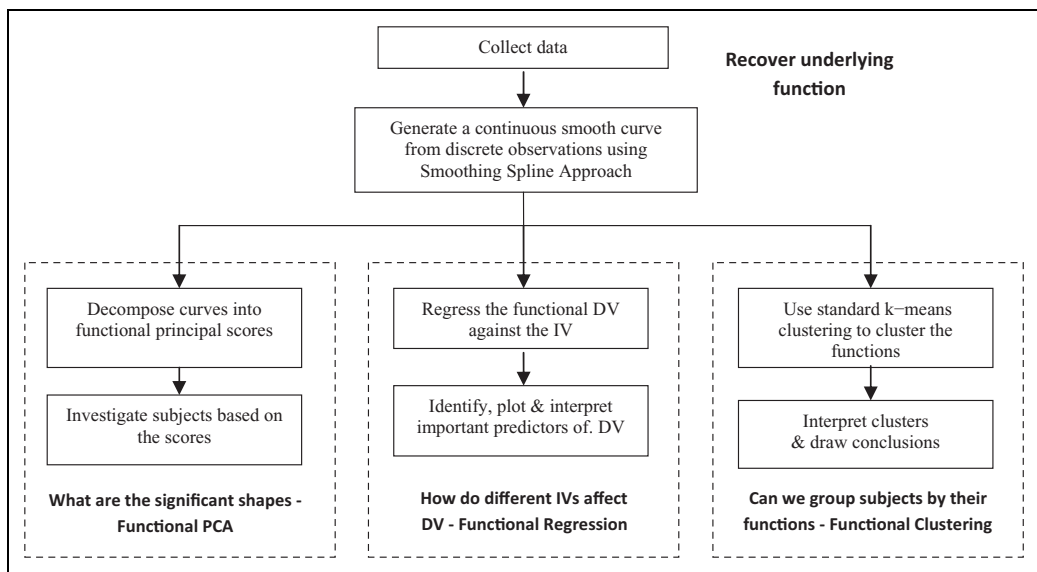
Figure 1. Flowchart of functional data analysis

Note: DV = dependent variable; IV = independent variable; PCA = principal component analysis.

constraint on the underlying function (Ramsay & Silverman, 2005). Therefore, we use a smoothing polynomial spline of order p of the following form:

$$f(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3, \ldots, + \beta_p t^p + \sum_{l=1}^{L} \beta_{pl} \left[(t - \tau_l)_+ \right]^p, \tag{1}$$

where τ_1, τ_2. τ_L is a set of L knots, t is time, and $u_+ = u I_{[u \geq 0]}$. The choice of L and p in Equation 1 determines the departure of the fitted function from a straight line with higher values resulting in a rougher f. The polynomial smoothing spline may result in a better fit of the observed data but typically tends to have a poorer recovery of the underlying trend with a tendency to overfit. To avoid this issue, a roughness penalty function (PEN) of the following form is imposed to measure the degree of departure from the straight line (Reddy & Dass, 2006):

$$\text{PEN}_m = \int \left[D^m f(t) \right]^2 dt, \tag{2}$$

where $D^m f$, $m = 1, 2, 3$., is the mth derivative of the function f. In other words, the value of PEN is high when the data points are highly nonlinear and the polynomial function (Equation 1) fits the data well and is low when the data points are linear. The goal is to find a function $f^{(j)}$ that minimizes the penalized residual sum of squares:

$$\text{PENSS}_{\lambda,m}^{(j)} = \sum_{i}^{n} \left(y_i^{(j)} - f^{(j)}(t_i) \right)^2 + \lambda \times \text{PEN}_m^{(j)}, \tag{3}$$

where $y_i^{(j)}$ represents the observed data j for organization i and $f^{(j)}(t_i)$ represents the corresponding functional value obtained from the smoothed spline. The smoothing parameter λ provides the trade-off between fit $[(y_i^{(j)} - f^{(j)}(t_i))^2]$ and variability of the function (roughness) as measured by PEN_m. The recovered function can now be further analyzed to compute its higher order derivatives (i.e., first-order derivative – velocity; second-order derivative – acceleration). However, studying

derivatives is only possible when the function is uncovered using a polynomial of order 3 or more, thus requiring at least four time points in the data. We used the b-spline module developed by Ramsay (2003) for minimizing $\text{PENSS}_{\lambda,m}^{(j)}$. The following algorithm performs the preceding process in R:

Step 1: Call the R library *fda*.
Step 2: Set the number of knots for spline fitting. This number depends on the number of observations for that particular subject.
Step 3: Run three sets of statements including *create.bspline.basis*, *fdPar*, and *smooth.basis*. *Create.bspline.basis* constructs the functional data object for the subject based on the number of knots, order of the basis functions, and the time points. *fdPar* generates another object by taking the output of the above statement, along with the smoothing parameters Lambda (roughness penalty factor) and Lfd (linear differential operator). Finally, *smooth.basis* smoothes the data using the roughness penalty (corresponding to Equation 3) and generates the function.
Step 4: Use the statement *deriv.fd* to calculate the first (velocity) and second (acceleration) derivatives of the function. See the code in R presented in the appendix for more information.

To illustrate the preceding process, consider performance data (ROA) of a firm shown in Figure 2a. We follow the preceding steps to discover the underlying function shown with a solid line in Figure 2b. Apart from recovering the underlying functions for the observed data, smoothing also helps in eliminating noise in the data (Foutz & Jank, 2010), and the resulting functions can be used to explore higher order views of the dynamics of the observed data, most commonly velocity (the first derivative of the discovered function) and acceleration (the second derivative of the discovered function), as shown in Figure 2c and 2d, respectively. Velocity measures the rate of change in the function over time, and acceleration indicates the rate of change in velocity over time.

Functional Principal Components Analysis

After recovering the underlying functions from the observed data, researchers may analyze these curves to extract the common factors that are present across all units of analysis (in our case, common trends across organizations). For example, we may want to know how many common trends are present in the performance of all firms and to identify them. This is done using fPCA. fPCA is derived from ordinary principal components analysis such that they have the same underlying concepts and goals (Foutz & Jank, 2010; James et al., 2000). Consider a set of observed data, say, $z_1, z_2, \ldots, z_n$, where each z is expressed in a p-dimensional data vector $z_i = (z_{i1}, z_{2i}, \ldots, z_{ip})^T$. For an ordinary PCA, the goal is to find a projection for $z_1, z_2, \ldots, z_n$ in a new space such that the variance along each component of the new space is maximum and orthogonal. Therefore, ordinary PCA finds a principal component (PC) vector $e_1 = (e_{11}, e_{12}, \ldots, e_{1p})^T$ for which the principal component scores (PCS) for p PCs are computed as

$$S_{ip} = \sum_j e_{pj} z_{ij} = e_p^T z_i \tag{4}$$

such that it maximizes $\sum_i S_{ip}^2$ to

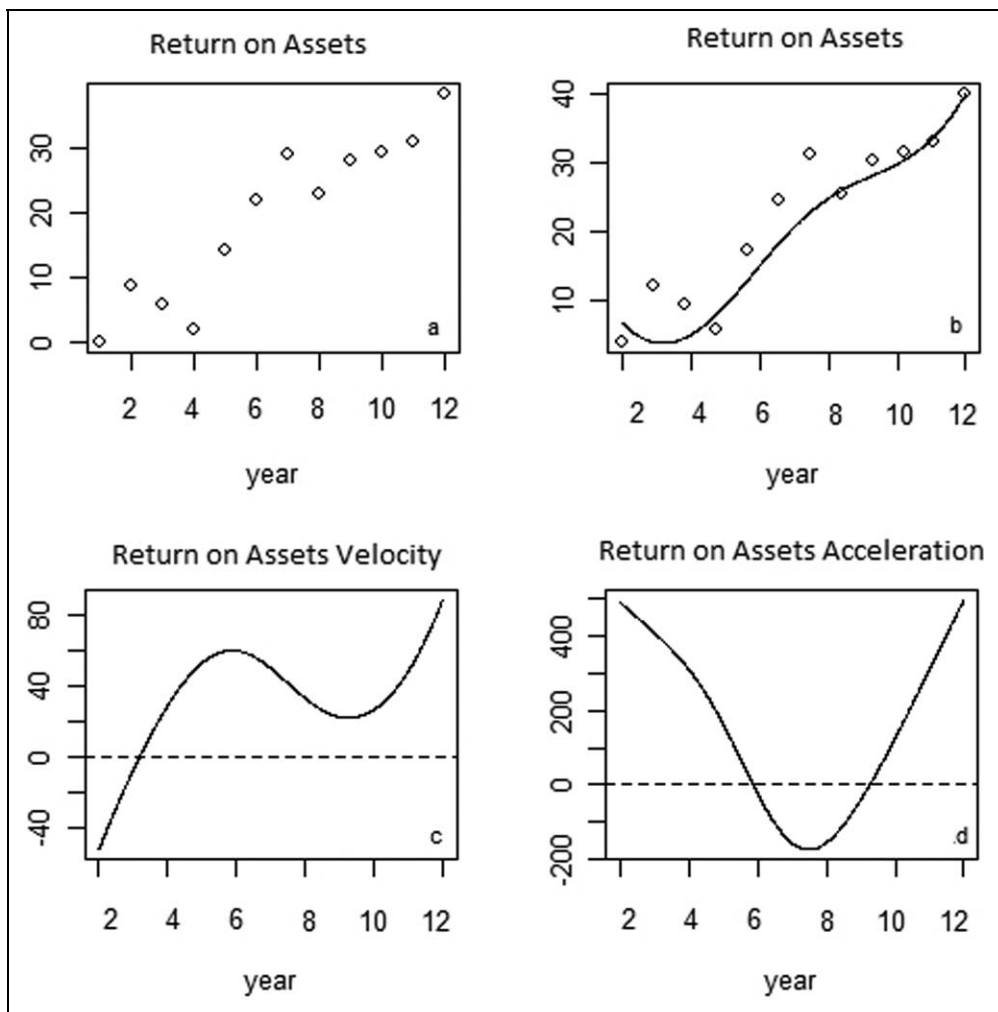$$\sum_j e_{pj}^2 = \left\| e_p \right\|^2 = 1. \tag{5}$$

Figure 2. Underlying function, velocity, and acceleration of return on assets of a firm

In the context of fPCA, the above procedure is repeated, except now the z is a set of continuous curves instead of discrete values. Consider a set of curves $z_1(s), z_2(s), \ldots, z_n(s)$ that is recovered from the previous step. Here, fPCA will find the corresponding set of PC curves $e_i(s)$ such that it maximizes the variance along each component and is orthogonal to each other. Therefore, for the each of the p principal components, the PCS is computed as

$$S_{ip} = \int e_p(s)z_i(s)ds, \tag{6}$$

such that it maximizes $\sum_i S_{ip}^2$ to

$$\int e_p^2 = \|e_p\|^2 = 1 \tag{7}$$

In other words, functional PCA is similar to ordinary PCA, except the modifications to adapt to functional data. The cutoff for selecting variance explained depends on the researcher. As a rule of

thumb, considering factors explaining 90% of the variance may be appropriate (James et al., 2000), although the researcher may decide, as with traditional PCA, that a lower threshold of variance explained is more sufficient for the research question. In R, we run a *princomp* statement on the function generated by the *smooth.basis* statement and identify the important factors. Refer to the appendix for the representative code.

Functional Regression

We can also investigate the relationship between predictor variables and response functions and determine how their relationship changes over time. For example, we may ask how firm size affects firm performance. Moreover, as both these measures change over time, we would also like to capture the dynamics of their relationships using functional regression. Unlike standard regression models where predictor and explanatory variables are scalars or vectors, functional regression allows these variables to take on a functional form. Therefore, the response variable is the recovered function of the observed data, and it is regressed on a set of functional predictors whose effects are the focus of the study. The estimation process is repeated across the sample period, allowing the researchers to understand the effect of the predictors on the response variable(s) over time. As Ramsay and Silverman (2005) point out, this is achieved by estimating $\beta(t_i)$ for a finite number of points in time t and constructing a continuous parameter curve by simply interpolating between the estimated values $\hat{\beta}(t_1), \ldots, \hat{\beta}(t_n)$. Therefore, the underlying model of the functional regression for the response function $y(s)$ is

$$y(s)_t = \beta_{0t} + \sum_{k=1}^{K} \beta_{kt} x(s)_{kt} + e. \tag{8}$$

If only two time points are observed at the subject level, or if there are static predictors in the data set, the underlying function will be linear and may not be enough to demonstrate changes in relationships over time. However, prior research using FDA presents an innovative approach to deal with static predictors (e.g., Dass et al., 2011; Wang et al., 2008), that is, *transforming the static variables into time-varying predictors* by considering each static variable's impact on the dependent variable. This is done by fitting a functional regression model of the dependent variable on each of the static predictors and then using their resulting time-varying estimated coefficient as a time-varying predictor (Dass et al., 2011, p. 1263).

In R, we run an *lm* statement to investigate the relationship between the function generated by the *smooth.basis* statement and the covariates in the model. See the appendix for sample code.

Functional Clustering

A third type of analysis that is often conducted using functional data is the grouping of units of analysis based on their underlying response functions (James & Sugar, 2003). For example, we can segment firms in our sample based on the dynamics of the firm performance functions. A k-means clustering is performed on the response curves, and the resulting membership information of the subjects is considered for grouping. For a set of response curves $z_1(s), z_2(s), \ldots, z_n(s)$, k-means clustering algorithm partitions n subjects into k sets $H = \{h_1, h_2, \ldots, h_K\}$ such that it minimizes the within-cluster sum of squares of the partitions; that is,

$$\arg\min \sum_{i=1}^{k} \sum_{z(s)_j \in H_i} \left\| z_j(s) - \mu_i \right\|^2. \tag{9}$$

This process differs from standard *k*-means clustering algorithm as it is performed on functional rather than discrete data. Therefore, in R, we run a *kmeans* statement on the function generated by the *smooth.basis* statement and identify the number of clusters based on reduction in the within-cluster dissimilarity. The appendix provides the representative code.

An Investigation of Firm Performance

To introduce FDA to organizational research, we explore trends and dynamics in financial performance using three common measures of performance: return on assets (ROA), Tobin's Q, and shareholder return. We provide a brief literature review in the following section on organizational performance measures and discuss the assumptions and expectations in using these metrics.

Firm financial performance is the primary outcome of interest in strategic management research, yet performance itself has been problematic as a construct (Boyd, Gove, & Hitt, 2005; Richard, Devinney, Yip, & Johnson, 2009). Numerous measures and approaches for estimating and describing firm performance exist. In a recent review, Richard and colleagues (2009) identify 213 papers appearing in top management journals between 2005 and 2007 (representing 29% of all articles published in these outlets) that model firm performance as a variable of interest. They find that 53% of the studies use an objective accounting-based measure (e.g., ROA), 17% use an objective financial market measure (e.g., return to shareholders), and 11% use Tobin's Q or other mixed measure.

The vast majority of managerial studies on performance use a single accounting or financial market measure. Given requirements of filings for publicly traded firms and thus the availability of accounting data, measures such as return on assets or equity are among the most well-established and accepted metrics for firm performance. However, they can also be "distorted by accounting policies, human error, and deception" and "emphasize historic activity over future performance" (Richard et al., 2009, pp. 727-728; see also Keats, 1988). Much finance and economics research relies on shareholder return as a more forward-looking metric of firm performance, reflecting market expectations for future performance, but a measure that is likewise limited. Stock returns are influenced by overall market volatility and biases (e.g., analysts, institutional owners, the media) and are able to reflect only the entire organization, as shareholder returns cannot be allocated by divisions or products that may be responsible for creating the stock's overall value in the market (Richard et al., 2009).

Given the limitations of unidimensional approaches, scholars have developed mixed measures of firm performance to reflect both accounting and financial market information. Among the most popular hybrid metrics is Tobin's Q, which compares the market value of assets to their replacement cost or book value. Miller (2004) and McGahan (1999) depict Tobin's Q as a forward-looking performance metric that adjusts for risk, whereas Lang and Stulz (1994) describe it as a longer term expectation of market valuation. Using the book value of assets as the denominator restricts the ratio to reflect historical rather than current replacement costs and also raises concerns that the measure fails to account for important but intangible organizational assets. Yet Tobin's Q remains the most prevalent mixed measure of performance to capture the extent to which a particular firm is under- or overvalued at a given time.

Understanding the dynamics and implications of the most common operationalizations of firm performance is critically important to management scholars, given that "organizational performance is the ultimate dependent variable of interest for researchers concerned with just about any area of management" (Richard et al., 2009, p. 719). We introduce our data set in the next section, including sample and measures, and then describe an investigation of firm performance over a 17-year period using functional data analyses. Although not exhaustive, the methods described here introduce several important functional techniques and illustrate the potential to apply FDA to the realm of organizational science.

Data

Our data set follows the 1,000 largest U.S. firms in terms of revenues from 1992 through 2008. Data from COMPUSTAT and CRSP are used to track the financial performance of each company across our sample period. We follow recent convention (e.g., Hawawini, Subramanian, & Verdin, 2003) in identifying outlying values that may skew our results. Firms that were significant outliers (i.e., more than three standard deviations from mean values for any of the three primary measures) for our performance measures were deleted from the sample. Thus, the final sample consists of 876 firms and 10,490 firm-year observations.

Measures

Our outcome of interest is financial performance; thus, we collect data on a number of traditional performance measures, including *shareholder return* (total return at fiscal year-end), Tobin's *Q* (compares the market value of assets to their replacement cost or book value; see Bertrand & Schoar, 2003), and *return on assets* (ratio of net operating profit to total assets). Although our demonstration focuses on the dynamics of firm performance, specifically across measures, the techniques also reveal how performance covaries with other firm characteristics, such as firm size. For our sample firms, we collect three measures of firm size: *assets*, *income*, and *total sales*.

Analysis and Results

In this article, we explore three measures of firm performance and investigate (1) how these measures vary across firms over time, (2) common trends or factors across performance of all firms, (3) the effect of various measures of firm size on these performance measures, and, finally, (4) clusters of firms present and their characteristics based on the dynamics of performance measures (using Tobin's *Q* for illustration).

   Discovering underlying functions: estimating functional forms. For all firms, we normalize time between 0 and 1 for ease in interpreting the results. This is done by dividing each time point by the total number of possible data points (Bapna et al., 2008; Reddy & Dass, 2006). This approach brings all firms into a common time frame and allows us to identify which firm data, if any, are missing at which time point. FDA can also be done without normalizing the data. Figure 2 illustrates the ROA function of a firm estimated without normalizing the data. As a first step toward analyzing firm performance dynamics, we discover the underlying functions for each firm by estimating Equation 1. We also estimate the rate of change of firm performance (velocity) and rate of change of firm performance velocity (acceleration) to examine the higher degree dynamics. These results are presented in Figure 3. As FDA is based on smoothing, and selection of smoothing parameters can significantly affect the outcome, we perform a sensitivity test (Ramsay & Silverman, 2005; Wang et al., 2008) to find the most appropriate knot and penalty parameter for the given data set. We considered multiple values of p (2, 3, 4, 5, and 6) and λ (10 different values between 0.001 and 100) and found that the model fit is insensitive to the different values of p and λ. However, since root mean square error (RMSE) for the model was lowest for $p = 4$ and $\lambda = 0.01$, we use these values to discover the underlying functions.

   As one would expect, these plots show high levels of heterogeneity across the firms in terms of their performance. For example, examining the plot of shareholder return as raw data (upper right figure), we find that some firms have steady growth, some have steady decline, and others have little change in returns over time. We can take a granular look at the dynamics of ROA with its velocity and acceleration plots and find that performance of some firms changes dramatically in the early
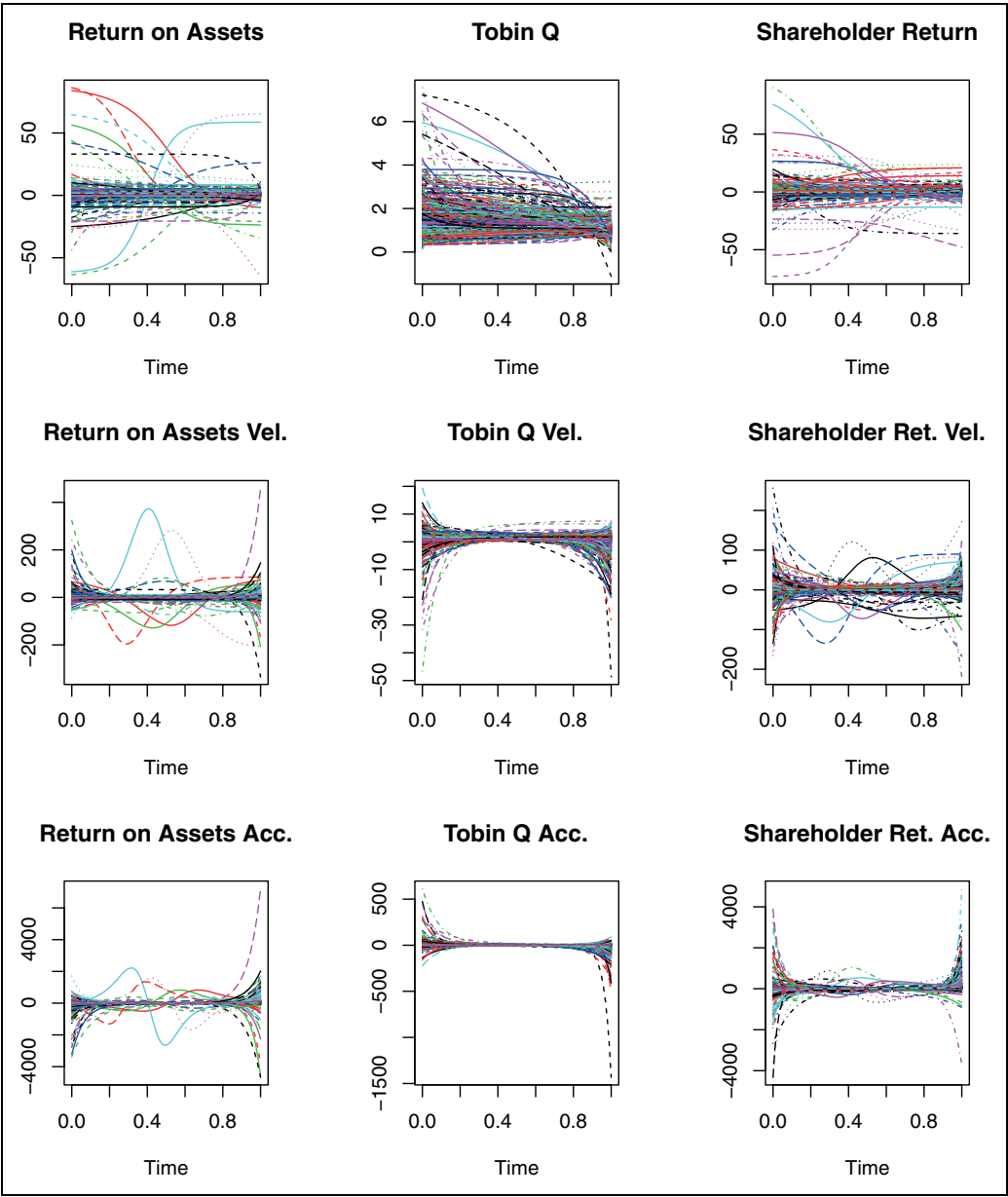
Figure 3. Return on assets, Tobin's *Q*, and shareholder return dynamics across firms

years, some during the middle, and others more recently, in later years of the sample period. Similar heterogeneity across firms is also observed for shareholder return and Tobin's *Q*.

One way to investigate how these performance measures and their dynamics evolve is by computing and plotting average values of the estimates (Figure 4). Examining these plots, we find that, on average, ROA and shareholder return increase over time, whereas Tobin's *Q* declines with time. Given that we track the 1,000 largest U.S. firms, we can see the macroeconomic effects in our sample from 1992 through 2008. Although both ROA and shareholder return have positive growth over the period, their rates of change differ. For example, the velocity and acceleration plot of ROA
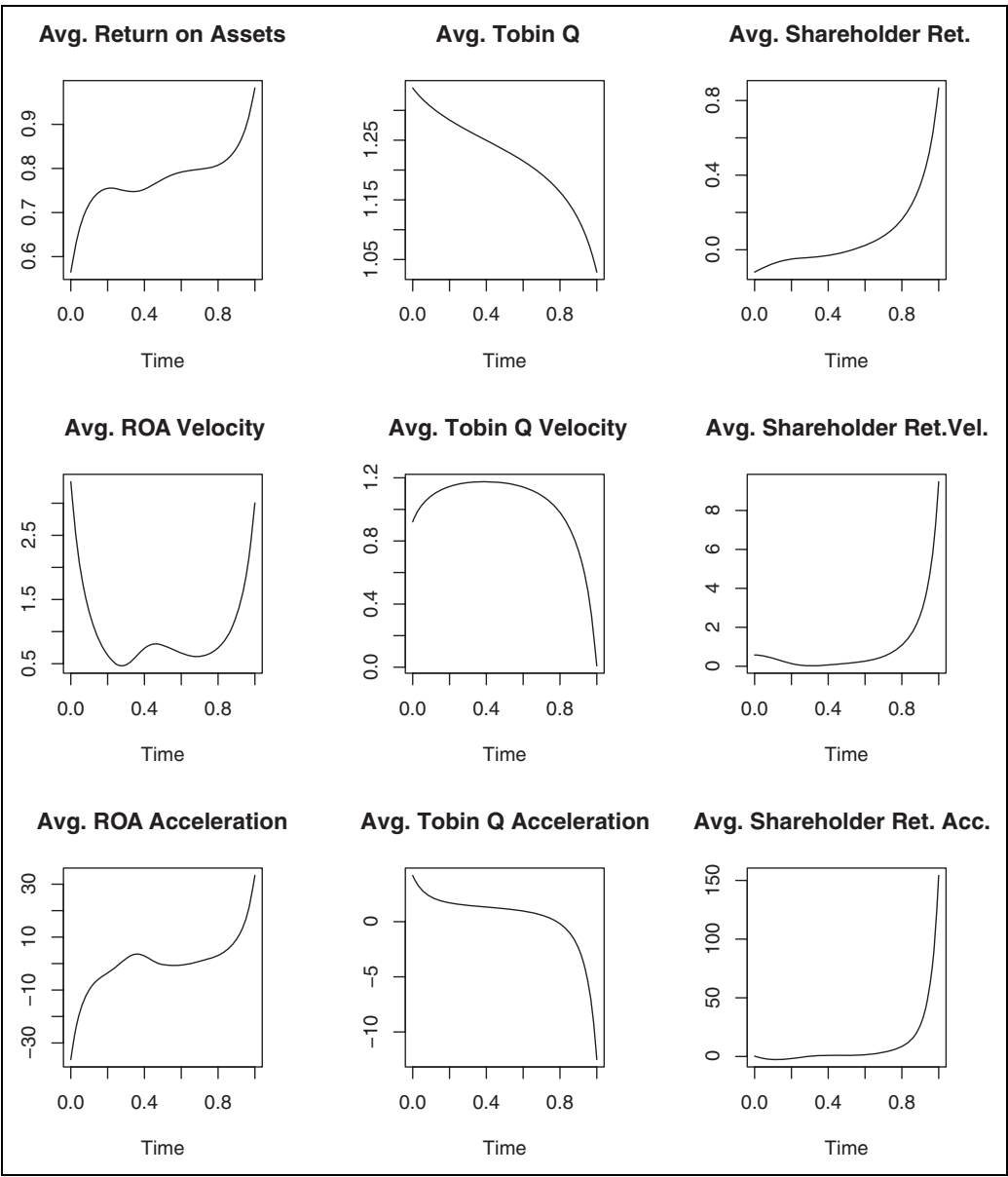
Figure 4. Average return on assets (ROA), Tobin's *Q*, and shareholder return dynamics across firms

suggests that firms typically have large changes at the early and later periods, with ROA growth decelerating at the early period and accelerating at the later period. On the other hand, shareholder return is typically stable until the later period, where it starts to enjoy acceleration. So the nature of ROA and shareholder return being backward versus forward looking, respectively, is confirmed in modeling the functional experience of the largest-revenue firms over 17 years, with the greatest increases in wealth occurring in the final time period (roughly 2004-2008). Similarly, as would be expected given the nature of the underlying data, Tobin's *Q* maintains a steady decline (limited change in the velocity plot) until the later period, where it suffers deceleration. Since Tobin's *Q* reflects the replacement value of assets, it is intuitive that market-to-book value would decrease with
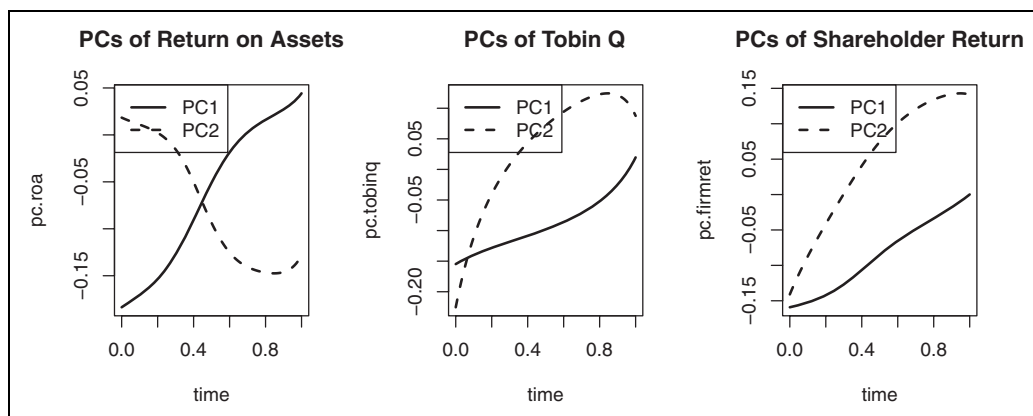
Figure 5. Functional principal component (PC) analysis plots of return on assets (ROA), Tobin's *Q*, and shareholder return

depreciation over time. The functional forms that emerge in the data help us explore and understand the dynamics and evolution in firm performance.

Generating typical shapes: functional principal component analysis. Functional plots of the performance measures provide an overview of how firms differ and of the patterns in their dynamics. Average function plots only indicate the average dynamics of the measures but do not reveal the typical shapes that are common across firms. To explore this question, we perform fPCA on these measures (Equations 6 and 7). In fPCA, identifying the number of principal components is particularly important as the fit of the different components is not independent, and choosing too many components can lead to overfitting. For ROA, the first PC explained approximately 66% of the variability, and the second PC explained around 29% of the variability. For Tobin's *Q*, the first PC explained 79% variance, and the second PC explained 16% variance, and for shareholder return, the first PC explained 77% and the second PC explained 18% of the variance. Since two PCs capture most of the variance, with a third component adding very marginally, we only select two and neglect the remaining components to avoid overfitting. We further performed a validation test suggested by James et al. (2000), which involves calculating the log likelihood for the reduced rank method by varying the number of possible PCs (p). Given that the fitting algorithm converges to the global maximum, the log likelihood is expected to increase with the number of PCs, but the increase should stabilize when the optimal rank is obtained. For all three measures, we found that the log likelihood value levels off after $p = 2$ following a significant jump from $p = 0$ to $p = 1$ and from $p = 1$ to $p = 2$. Thus, we considered all three measures to have two principal components. We plot these components (Figure 5) and investigate what they represent. The PC plots of ROA show two opposing components, with one representing an increase in ROA over time, whereas the other represents a reduction. If a firm has positive ROA, it will load heavily on the increasing PC and vice versa. In comparison, PCs for Tobin's *Q* show a negative PC for most of the time, suggesting a decreasing trend similar to what the average plot suggested and aligned with the calculation of the metric itself. PCs for shareholder return are positive and increasing, thus indicating the overall trend and macro-effects across firm-years that we observed earlier.

Investigating time-varying relationships: functional regression. After examining the common curves of the three performance measures, we examine how other firm characteristics, including sales, net income, and assets, affect these performance outcomes. To do so, we recover the functions of these
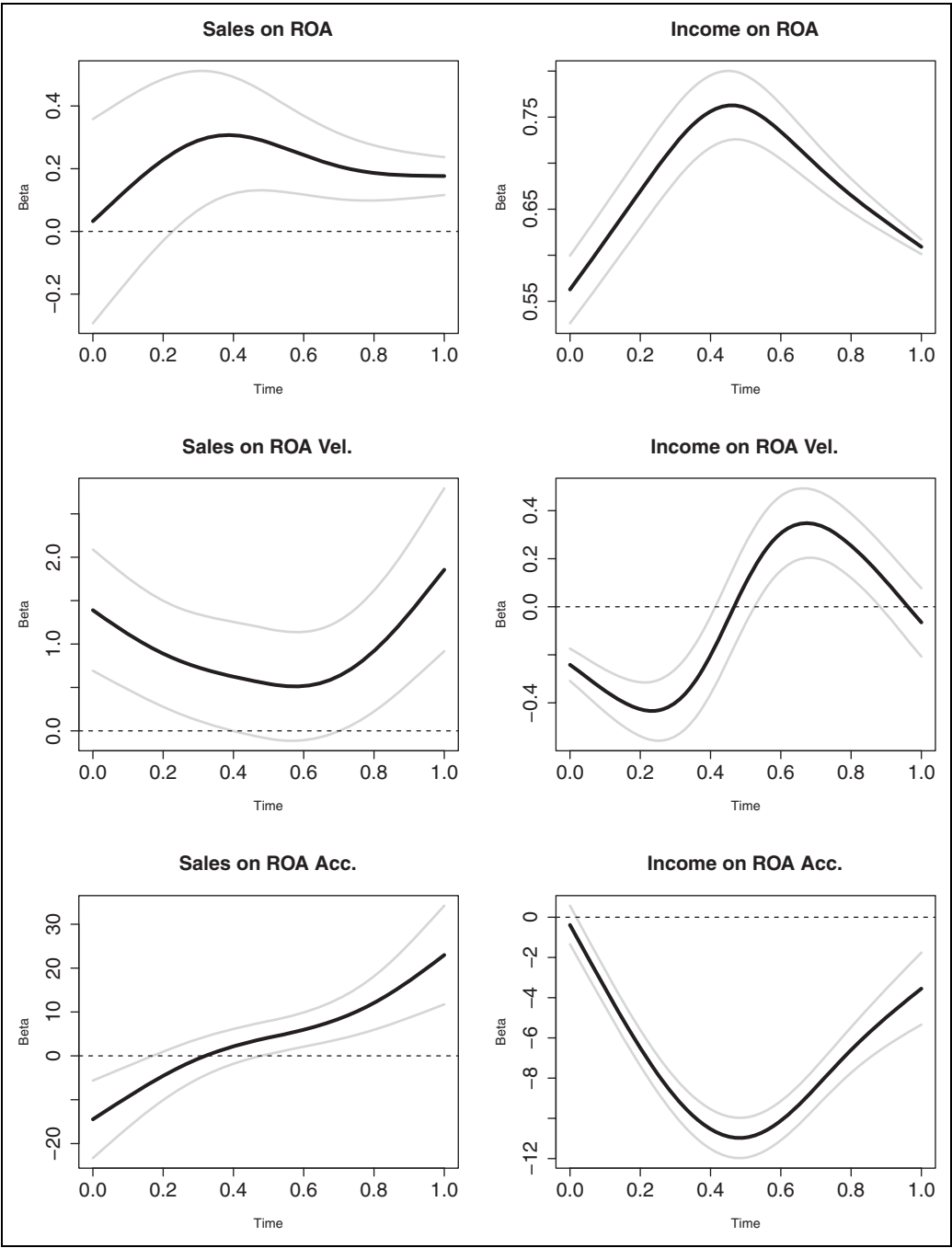
Figure 6. Coefficient plots for return on assets (ROA)

factors as above and conduct functional regression (Equation 7) using functions of ROA, Tobin's Q, and shareholder return as dependent variables and the above firm size measures as independent variables. The process estimates the effects for each time period and then plots the coefficients of each independent variable with respect to time. The confidence intervals for the coefficients at
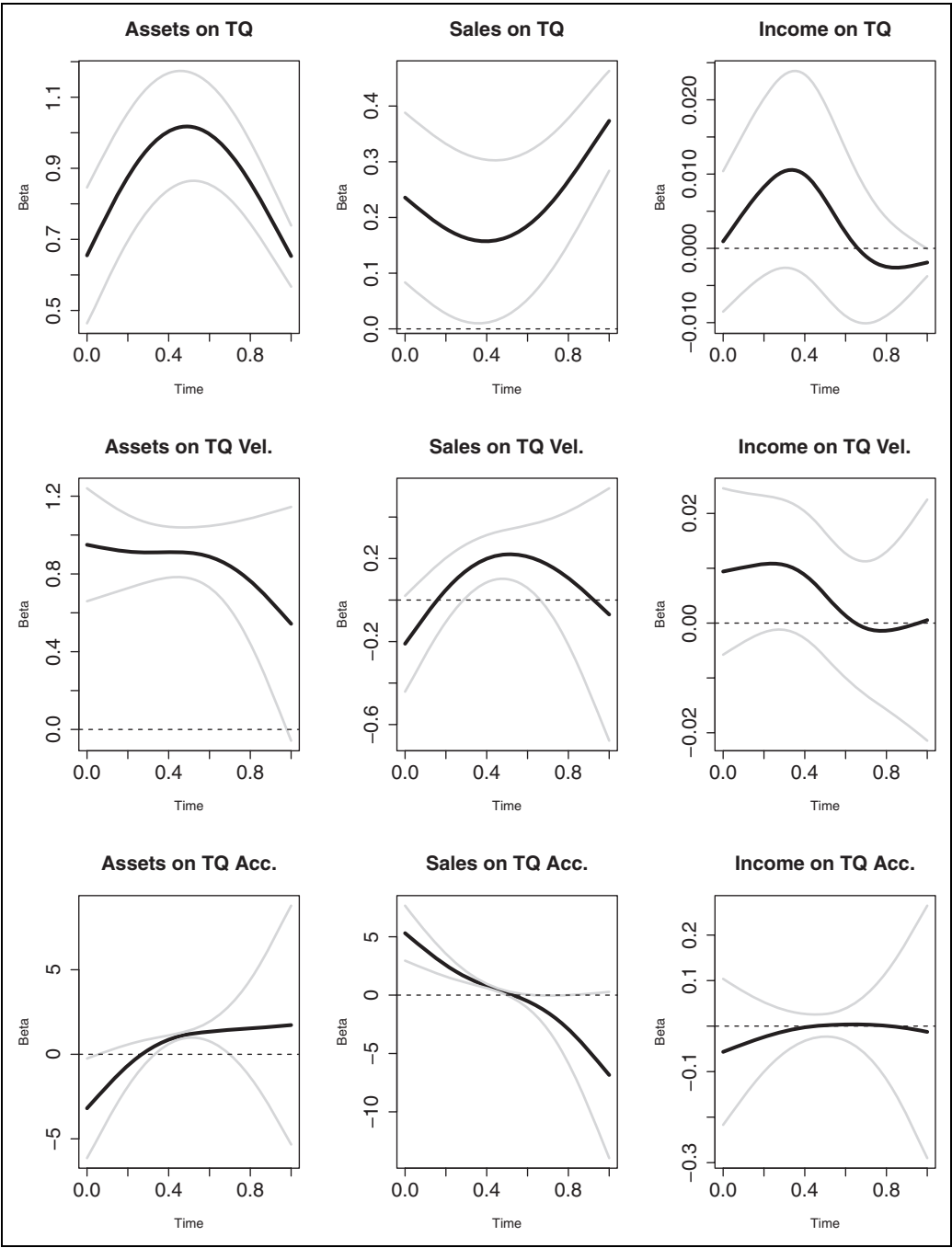
Figure 7. Coefficient plots for Tobin's *Q* (TQ)

95% are also plotted. If the confidence band includes the zero line at any time point, the effect of the corresponding factor is not significant. We estimate the model for the performance measurements and their dynamics (velocity and acceleration).
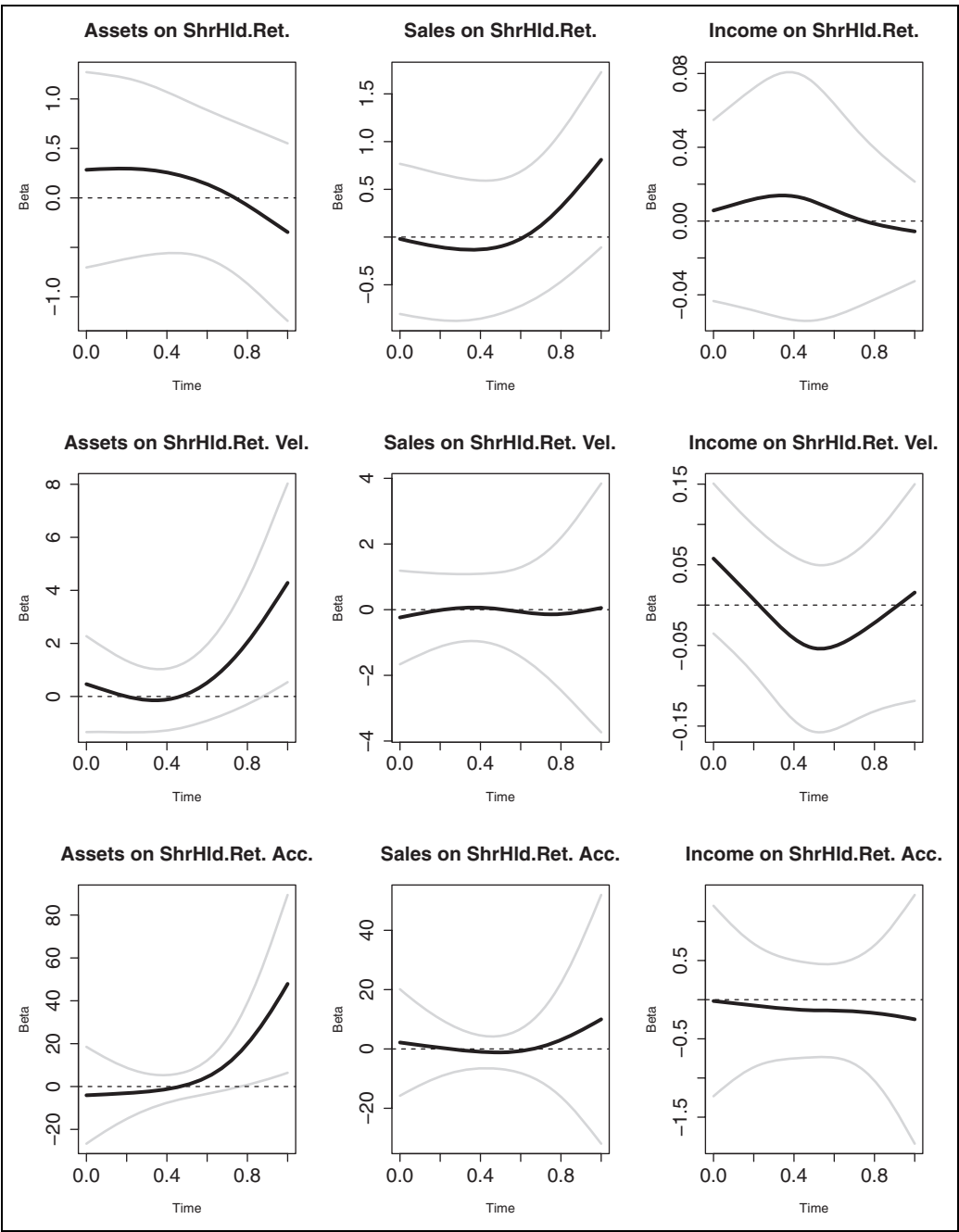
Figure 8. Coefficient plots for shareholder return

Results for ROA are shown in Figure 6. In this case, we examine how sales affect ROA over the period. We find that sales have a nonsignificant effect on ROA in the early years but a significant and positive effect thereafter. In other words, after the initial period, increasing sales lead to higher ROA. On examining the effect of sales on ROA velocity, we find positive and significant effects throughout the period, with the effect declining slightly until the midpoint in time and then
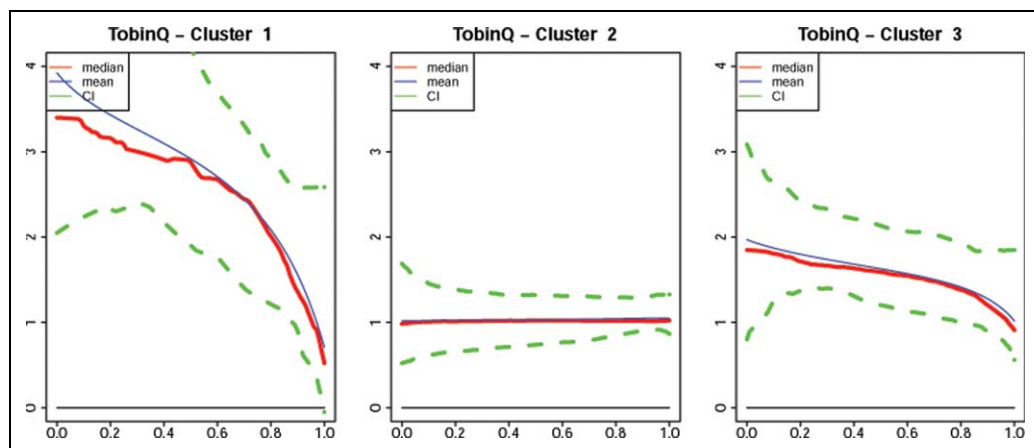
Figure 9. Tobin's Q plot across three firm clusters
Note: CI = confidence interval.

increasing again in the later period. This indicates that increasing sales slow the increase in ROA as the firm operates from the early period to the midpoint but then quicken the ROA increase at the later period. This trend is also reflected on the plot of effect of sales on ROA acceleration. On the other hand, income has an inverted U-shaped, positive, and significant effect on ROA with the highest effect during the middle years. The plot for ROA velocity also reflects the same trend with negative velocity from the early to the middle years, and then positive velocity from mid to late years. The plot for ROA acceleration reiterates the above finding but shows that income has a negative effect on ROA acceleration throughout the time period.

It is important to note that functional regression analysis not only uncovers the effect size but also indicates how the effect changes over time. Thus, functional regression presents a simple yet efficient approach for investigating dynamics of firm measures and their relationship to one another. For Tobin's Q and shareholder return, we considered sales, income, and assets as our independent variables. Results (Figure 7) show that assets and sales have a significant effect on Tobin's Q, with sales having a U-shaped and assets having an inverted U-shaped effect over time. On examining their effect on Tobin's Q velocity, we find that assets have a significant and positive effect on velocity throughout the period, that sales have a significant effect only during the mid-period, and that income has no significant effect. Interestingly, the effect of assets on Tobin's Q acceleration is only significant through the middle period, and sales have a positive effect between the beginning and midpoint of the sample period. On examining the effects on shareholder return (Figure 8), we find that none of the firm size factors has a significant effect on the dependent variable.

The ability to model changes in financial performance as driven by changes in other firm characteristics is exciting, given the sizable body of management research predicting firm performance. Although previous studies use average size or attempt to decompose accounting profitability or Tobin's Q as performance measures (e.g., McGahan, 1999; McGahan & Porter, 2002), we still lack understanding of the interrelationships between changes among firm characteristics across time. Even in our demonstration of functional regression, when looking at changes in sales predicting changes in Tobin's Q, we uncover interesting evidence that although sales have a positive effect on firm performance generally, the effect declines over time. Our results hint at the novelty and promise of functional regression in understanding relationships among organizational phenomena across time.

Table 1. Estimation Using a Hierarchical Linear Model

Fixed Effect	Coefficient	Standard Error	t Ratio	p Value
For INTRCPT1, π_0				
INTRCPT2, β_{00}	0.057977	0.003714	15.612	<.001
RET, β_{01}	0.064148	0.008495	7.551	<.001
SALES, β_{02}	0.000000	0.000000	1.997	.046
For TIME slope, π_1				
INTRCPT2, β_{10}	−0.214515	0.028179	−7.613	<.001
RET, β_{11}	−0.405986	0.064611	−6.284	<.001
SALES, β_{12}	0.000000	0.000001	0.591	.555
For TIME2 slope, π_2				
INTRCPT2, β_{20}	0.556213	0.064463	8.628	<.001
RET, β_{21}	0.651881	0.147448	4.421	<.001
SALES, β_{22}	−0.000002	0.000002	−0.965	.334
For TIME3 slope, π_3				
INTRCPT2, β_{30}	−0.378652	0.042647	−8.879	<.001
RET, β_{31}	−0.316994	0.097463	−3.252	.001
SALES, β_{32}	0.000001	0.000001	0.981	.327

Grouping similar firms: functional clustering. Finally, we explore the notion of grouping firms based on the dynamics of their performance measures. For simplicity, we consider only Tobin's Q as the clustering criterion. We use a standard k-means algorithm on the Tobin's Q function (Equation 8) and use a between-segment distances (BSD) measure to identify the number of clusters. In the BSD approach, we first compute the reduction in the within-cluster dissimilarity and then identify the cluster solution where the reduction is the largest. In our case, we found that firms can be grouped into three clusters based on their Tobin's Q values. As there are various criteria to determine the number of clusters, the decision on which criteria to select is at the discretion of the researcher, similar to standard cluster analysis using k-means. Although we used BSD in this research, other criteria can also be used. After identifying the clusters, we explore their characteristics. Since Tobin's Q was the grouping criterion, we investigate the differences across the clusters as illustrated in Figure 9. We find that the firms belonging to Cluster 1 ($n = 35$) have Tobin's Q that decreases over time. Firms belonging to Cluster 2 ($n = 656$) exhibit similar characteristics, but the slope of the decay is much flatter than that of Cluster 1. Finally, firms in Cluster 3 ($n = 185$) display a stable and near-constant value across time. We can further investigate cluster membership to compare and contrast groups within the sample and their characteristics. For example, Cluster 1 includes on average firms that employ the fewest employees (mean = 25,700) yet invest the most in research and development (R&D) (mean = \$347 million) and have the highest total risk (mean = 0.0269, standard deviations in daily returns over the fiscal year). Firms in Cluster 2 are medium-sized firms, with an average of 29,900 employees, \$117 million in R&D expenditures, and yet the lowest total risk (mean = 0.0125). On average, Cluster 3 includes the largest firms, averaging 36,660 employees, spending \$212 million on R&D annually and total risk of 0.0228. We present these sample characteristics by cluster membership to illustrate how, depending on the research question, functional clustering allows researchers to identify groups within the sample that follow similar paths or patterns of change.

Comparing FDA with HLM. We have suggested that FDA is suitable in certain situations where the functional form of subjects varies and considering one common structure is inefficient and that FDA has the ability to handle time-varying relationships among the variables. To show these two

advantages, we analyze ROA with shareholder return (RET) and sales (SALES) using HLM. Since we are not aware of the trajectory of change in the Level 1 model, we estimated five HLM models with different polynomial trajectories, including a no-change model, linear change (first-order) model, quadratic change (second-order) model, a cubic change (third-order) model, and a fourth-order change model. Polynomial models were used to make the Level 1 model similar to that of the function discovery process (Equation 1) in FDA. Next, we compared deviance statistics (Singer & Willett, 2003: 220) of these five models to find the most suitable model for analysis. We found the difference between the linear change model and no-change model to be 14.59, which exceeds the 0.01 critical value of χ^2 on 3 *df*, and thus, the linear change model is superior to the no-change model. Similarly, the difference in deviance statistics between the quadratic change and linear change model is 17.41, which is more than $\chi^2 = 13.28$ at the 0.01 critical value and with 4 *df*, and the difference in deviance between the quadratic change and cubic change is 93.18, which is more than $\chi^2 = 20.52$ at the 0.001 critical value and with 5 *df*. Since the deviance statistic increases with the fourth-order change model, we conclude that the cubic model is the most appropriate trajectory for our data. Therefore, we use the following model specification for the HLM analysis of our data:

Level 1 model:

$$ROA_{ti} = \pi_{0i} + \pi_{1i}{}^*(TIME_{ti}) + \pi_{2i}{}^*(TIME2_{ti}) + \pi_{3i}{}^*(TIME3_{ti}) + e_{ti} \tag{10}$$

Level 2 model:

$$\pi_{0i} = \beta_{00} + \beta_{01}{}^*(RET_i) + \beta_{02}{}^*(SALES_i) + r_{0i.} \tag{11}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}{}^*(RET_i) + \beta_{12}{}^*(SALES_i). \tag{12}$$

$$\pi_{2i} = \beta_{20} + \beta_{21}{}^*(RET_i) + \beta_{22}{}^*(SALES_i). \tag{13}$$

$$\pi_{3i} = \beta_{30} + \beta_{31}{}^*(RET_i) + \beta_{32}{}^*(SALES_i). \tag{14}$$

Mixed model:

$$\begin{aligned} ROA_{ti} = {}& \beta_{00} + \beta_{01}{}^*RET_i + \beta_{02}{}^*SALES_i + \beta_{10}{}^*TIME_{ti} \text{-} \beta_{11}{}^*RET_i{}^*TIME_{ti} \\ & + \beta_{12}{}^*SALES_i{}^*TIME_{ti} + \beta_{20}{}^*TIME2_{ti} + \beta_{21}{}^*RET_i{}^*TIME2_{ti} \\ & + \beta_{22}{}^*SALES_i{}^*TIME2_{ti} + \beta_{30}{}^*TIME3_{ti} + \beta_{31}{}^*RET_i{}^*TIME3_{ti} \\ & + \beta_{32}{}^*SALES_i{}^*TIME3_{ti} + r_{0i} + e_{ti}, \end{aligned} \tag{15}$$

where ROA = return on assets, TIME = time variable, TIME2 = square of time variable, TIME3 = cube of time variable, RET = shareholder return, and SALES = sales.

We estimate the preceding equations using the HLM 7 software, and the results are shown in Table 1. Although HLM estimates the overall effects of the covariates, it does not reveal the time-varying relationships between ROA and shareholder return and sales as examined through FDA. Moreover, HLM does not allow us to estimate the dynamic characteristics (i.e., velocity and acceleration) of ROA or to investigate the time-varying relationships between them and other covariates (Figure 6). Results obtained from HLM show significant effects of shareholder return and sales on the intercept, whereas only firm return is significant at the .05 level for the higher order time slopes.

To show that restricting the within-subject functional structure to one form may lead to inefficiency, we compare the errors of estimated ROA within-subject functions from both methods. In particular, we fit the Level 1 model shown in Equation 10 and the underlying function obtained through FDA as shown in Equation 1 and compare the error from the two approaches. We find that the error from the FDA approach (error = 0.043) is significantly smaller than that of HLM (error =
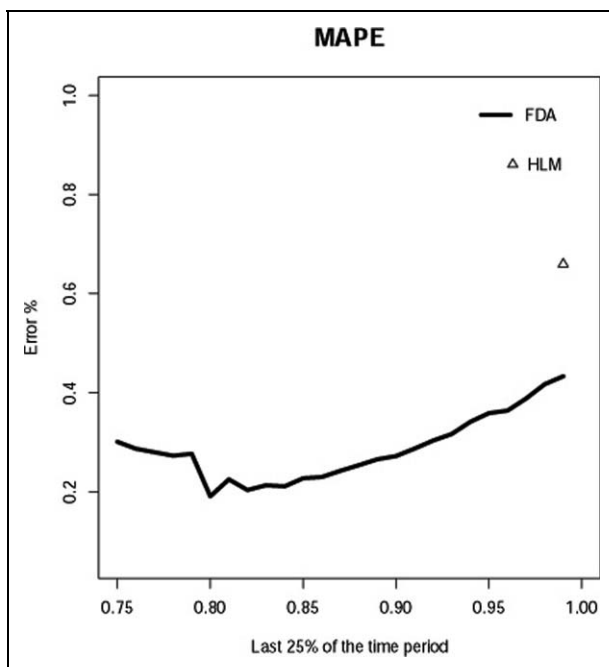
Figure 10. Mean absolute percentage error (MAPE) comparison between functional data analysis (FDA) and hierarchical linear modeling (HLM)

0.162) ($p < .05$), thus suggesting that fitting one common Level 1 model for all subjects is inefficient in this context. Since our context exhibits heterogeneity in the growth process, a random effect HLM model may account for some of the model misfit identified earlier. These results also highlight the advantages of using FDA, which include its ability to efficiently capture the underlying function of individual subjects, capture the common trends among them through functional principal components, and use these functions as clustering criteria in functional clustering.

Finally, to compare the predictive advantage of FDA over HLM, we develop an FDA-based dynamic forecaster (Dass et al., 2011) to predict ROA with sales and shareholder return and then compare it with an HLM-based forecaster. The FDA forecaster is called a dynamic forecaster as it uses the dynamics of ROA functions along with the predictor variables as the basis of forecasting and is able to dynamically predict the outcome variable in a later time period. In particular, it uses the following model for ROA at time t ($y(t)$) as given as

$$y(t) = \alpha + \sum_{i=1}^{2} \beta_i x_i(t) + \sum_{j=1}^{2} \gamma_j D^{(j)} y(t) + \sum_{l=1}^{L} \eta_l y(t-l) + \varepsilon(t), \tag{16}$$

where $x_1(t), x_2(t)$ is a set of time-varying predictors (i.e., sales and shareholder return), $D^{(j)} y(t)$ denotes the jth derivative of ROA at time t, and $y(t-l)$ is the lth ROA lag. Next, using Equation 16, the *h-step ahead* forecast, given information until time T, is estimated by

$$\tilde{y}(T+h|T) = \hat{\alpha} + \sum_{i=i}^{2} \hat{\beta}_i x_i(T+h|T) + \sum_{j=1}^{2} \hat{\gamma}_j \tilde{D}^{(j)} y(T+h|T) + \sum_{l=1}^{L} \hat{\eta}_l \tilde{y}(T+h-1|T). \tag{17}$$

The preceding functional forecaster is equivalent to an autoregressive (AR) model. For more information on the characteristics of the FDA forecaster and the model details, see Dass et al.

(2011). On the other hand, the HLM-based forecaster uses Equation 15 as the basis of forecasting and thus is similar to the static forecasters discussed in prior literature (Dass et al., 2011; Wang et al., 2008). The coefficients of the forecasters are first estimated using data from the training set (70% of sample, 613 firms) and then used to predict the values of ROA of firms in the validation set (30% of sample, 263 firms). The FDA forecaster predicts ROA function for each normalized time $t = 0.76$ to $t = 1.0$, using information available at time $t = 0.75$ and the time-varying lag value of ROA dynamics (Equation 17). The HLM-based forecaster is a static forecaster and thus only used to predict ROA at normalized time $t = 1.0$ for the 263 firms. Therefore, given their functional versus discrete approaches, the FDA and HLM forecasts yield differing output in terms of predicting a function over time versus a single point in time. We also compute mean absolute percentage error (MAPE) (Wang et al., 2008), illustrated in Figure 10, to compare their accuracy. We found that the forecaster based on HLM has an error of 66%, whereas the error from the FDA forecaster ranges from 19% to 43%, thus suggesting that the FDA-based forecaster has better accuracy than the HLM forecaster.

Promising FDA Applications in Organizational Research

Although our demonstration showcases three functional techniques to explore dynamics in firm performance measures across time, there are many promising FDA applications in organizational science research. Given the multiple levels inherent and complexity of relationships within and between organizations and their environments, methods that offer continuous rather than discrete approaches to statistical modeling get closer to the reality of the data themselves. We outline an agenda for future research that incorporates functional data techniques to enrich our understanding of both micro- and macro-organizational phenomena.

Organizational behavior research often relies on primary data collection, wherein survey or lab study participants provide feedback at various time points. Survey research typically collapses responses into windows of time to use traditional methods that rely on discrete data points. Functional techniques allow each participant's responses to be a function, so rather than discrete Time 1 versus Time 2 versus Time 3 observations, the timing and spatial dimensions across the entire study period can be captured. Incorporating research design to record exact response times (e.g., electronic time stamp), if participants are given 1 month to complete each of two questionnaires, with 2 months between surveys, then "Time 1" surveys submitted on Day 1 versus Day 30 can be captured functionally, and we can analyze differences (e.g., relationships with other covariates) between responses submitted by one participant on Day 1 and Day 120 relative to another participant responding on Day 30 and Day 90. These replies represent different intervals in response to an intervention, yet most survey research collapses replies by sampling period (e.g., Time 1, Time 2) as a matter of practicality and convenience. Meaningful information is lost in aggregating the data to broad categories and subsequently in the structural equation models using these inputs as equivalent time intervals and equally spaced observations. We offer this generic example to illustrate how functional techniques provide micro-researchers the opportunity to get closer to the realities of their data and depict responses as a function of time and space, which may have powerful implications both theoretically and empirically.

Functional analyses also offer scholars the opportunity to design studies that minimize common method biases sourcing from affect change, transient mood, or other artifacts of research collecting multiple responses across time (Podsakoff, MacKenzie, Jeong-Yeon, & Podsakoff, 2003). Organizational behavior researchers often use experience-sampling to study within-individual change and then model the multivariate dynamics of micro-phenomena such as motivation (Dalal & Hulin, 2008), job performance (Dalal, Lam, Weiss, Welch, & Hulin, 2009), and negative affect at work (Ilies, Johnson, Judge, & Kenney, 2011). We discuss the research design and modeling of recent published work to evidence the potential of FDA at capturing the dynamics and dimensionality

of micro-data. In one experience-sampling study, Judge, Scott, and Ilies (2006) collect daily responses over 3 weeks to explore intraindividual differences in job satisfaction and workplace deviance and suggest that further testing may reveal other associations and alternative representations of interest. Gathering responses over consecutive days or weeks, each participant's Time 1, Time 2, and so forth are not simultaneous but are treated as such in traditional models. A functional approach allows each subject to have its own functional model rather than restricting to a single Level 1 form for hierarchical modeling. Following Ilies and Judge's (2005) finding that "affect and goals are dynamically related within individuals in that they vary in synchrony across time" (p. 464), we suggest that functional approaches may be fruitful to explore trends and covariation in change among variables, within- and between-person changes, and potential moderation and mediation effects at multiple levels across time.

Macro-organizational scholarship also has much to gain from functional techniques. Multilevel methods are popular in research on strategic management; however, these efforts are still limited to structuring the data in a multivariate (i.e., set of individual numbers or values) rather than a functional way. Recognizing the experience by subject or company across time, FDA further separates variance by allowing curves to vary vertically (i.e., amplitude variation) as well as horizontally (i.e., phase variation), an important extension to understanding organizational observations as time-series data. FDA enriches the complexity of models and can separate the magnitude and direction of effects, as previously problematized in terms of the scope and pace of organizational change (e.g., Street & Gallupe, 2009). Studying the time-varying relationships between variables is particularly promising to macro-researchers, such as in the introductory example of entrepreneurship research (e.g., Holcomb et al., 2010). A recent study on international strategies and foreign direct investment argues the "need to examine FDI ownership structure as a complex and interrelated phenomenon" (Mani, Antia, & Rindfleisch, 2007, p. 865). The authors note more accurate and unique findings revealed by their adoption of longitudinal modeling, which still applies a consistent base model for their sample of 4,459 subsidiaries established by 858 firms across 38 countries in 9 years. We suggest that, like moving from traditional fixed effects to hierarchical estimations, freeing individual observations to have their own functional form allows researchers to further uncover the effects of unobserved heterogeneity and covariation in variables at multiple levels and across time.

As demonstrated in our firm performance example, FDA also allows for graphical depictions of the underlying functional data, especially higher order changes and patterns. The shape of derivative functions (i.e., velocity and acceleration) may reveal the influence of exogenous or interesting events undetected in plotted raw data and linear relationships. Macro-organizational scholars may find functional coefficient models useful to extend time-series or survival analyses (Fan & Zhang, 2008) or regression analyses to include functional predictors and functional outcomes. For example, researchers in social issues in management have long investigated the relationship between corporate social performance and financial performance (e.g., meta-analysis by Orlitzky, Schmidt, & Rynes, 2003), yet traditional methods do not fully capture how changes in social performance may be driving changes in financial performance or vice versa. As illustrated earlier, functional regression is particularly promising in allowing scholars to discover the dynamics of change among multiple organizational characteristics and outcomes over time.

Furthermore, functional methods offer the opportunity to build typologies of change over time by identifying functional components, recognizing breakpoints in derivative plots to classify life cycle stages, and functional clustering to identify categories or patterns of change. For example, Lynall, Golden, and Hillman's (2003) theory on the evolution of board composition and firm performance over time remains largely untested, perhaps due to the lack of methods capable of appropriately modeling changes and examining how firm characteristics covary as the firm evolves through its life cycle. Functional regression and clustering offer fruitful approaches to investigate changes and

stages of growth in board composition relative to a firm's evolving resource needs, as well as their interactions and ultimate relationship with firm performance.

Finally, although we illustrate three of the most common functional techniques, additional functional methods may be useful to managerial science. Previous work has used multivariate approaches such as canonical correlation and discriminant analyses to test congruence among business strategy, pay policy, and firm performance (e.g., Montemayor, 1996) and combinations of resources and capabilities as a source of competitive advantage (e.g., Carmeli & Tishler, 2004). Functional canonical correlation and other functional extensions of traditional methods can address extensive data or variable sets to explore these macro-research questions of strategic fit, compensation, and bundling relative to firm performance or, similarly, to explore micro-research questions involving person-organization fit, motivation, and personality relative to outcomes such as job satisfaction or workplace deviance. The possibilities for applying FDA within organizational scholarship are nearly as limitless as the research questions themselves, and we hope the ideas and topics suggested earlier are only the start of productive applications of functional methods.

Conclusion

In conclusion, we believe organizational scholars have much to gain by adopting functional techniques into their analytical toolkits. In particular, identifying the dynamics of change within and between units and over time is a common dilemma in organizational behavior and strategic management research, and FDA offers substantial advantages over traditional linear models through functional or hybrid research design. For example, functional regression can have functional inputs and outcomes that are either scalar or curves themselves. Past special issues of statistical journals provide additional support and explanation for functional topics (e.g., Davidian, Lin, & Wang, 2004; Gonzalez-Manteiga & Vieu, 2007). Further software support and sample code are available in Ramsay, Hooker, and Graves (2009), with links to code that can be downloaded. Fundamentally, FDA offers functional counterparts of well-known linear analyses, so we feel the methodological benefits far outweigh the limited learning curve and complexity associated with adopting these functional approaches.

We recognize that functional data analysis is still largely in its infancy and, as such, has some limitations in terms of both technical support and immaturity to resolve a number of more complex, functional research questions. Jank and Shmueli (2006) argue that FDA is vastly superior to static models with its ability to capture processes and dynamics over time. However, FDA is still evolving to deal with such challenges as concurrency of events and longitudinal changes in the functional objects themselves. In relation to our demonstration on firm performance measures, we see promise of functional techniques to answer recent calls for research that "examines triangulation using multiple measures, longitudinal data and alternative methodological formulations as methods of appropriately aligning research contexts with the measurement of organizational performance" (Richard et al., 2009, p. 718). As FDA controls for time-invariant errors, it reduces the influence of common method errors and corrects for contextual firm-specific fixed effects, thus moving beyond parametric approaches to more accurately address the dimensionality of organizational phenomena. Overall, we encourage managerial scholars to follow the momentum and advances in other disciplines recognizing functional data analysis as a promising, novel methodology to enrich our empirical and theoretical understanding of the research questions we ask and their underlying constructs, relationships, research design, and outcomes. We look forward to the broader adoption of FDA by micro- and macro-organizational scholars to enhance our understanding of managerial science.

Appendix

Sample R Code for Functional Data Analysis

```
rm(list = ls(all = TRUE))
#############################################
# Reading the required library files
library(splines)
library(pspline)
library(fda)
library(stats)
library(dlnm)
require(graphics)

#############################################
#Setting the seed.
set.seed(100);

#############################################
# Simulating Data for Illustration

times<-matrix(nrow=101, ncol=1);times <- seq(0,1,0.1)
nl.data<- arima.sim(list(order = c(1,1,0), ar = 0.1), n = 10);
a1 <- matrix(c(times));a2 <- matrix(c(nl.data))
orm <- cbind(a1, a2,1);id <- orm[,3];yvar <-orm[,2];xvar <- orm[,1];
#############################################
# Identifying the required parameters

pointer <- 1;k <- 1;
ormid <- id;
xtime <- xvar;
numdata <- length(yvar);
numid <- length(unique(ormid));
grid <- seq(0,1,0.01);
norder <- 6; # determined by the researcher
Lfd <- int2Lfd(4); # 4 for B-splines or 2 for monosplines
lambda <- 0.05; # determined by the researcher
nbasis <- length(grid) +norder-2; #This is the no. of Knots

#############################################
# Estimating the underlying function

bidbasis <- create.bspline.basis((0:1), nbasis, norder, grid);
bidfdPar <- fdPar(bidbasis, Lfd, lambda);
coefMat <- matrix(0);
betaMat <-matrix(0);
rel_yvar<-matrix(0);
resultfunc <- matrix(0, nrow=101, ncol=1);
firm <- ormid[pointer];
first <- pointer;
for(i in pointer: numdata){
   if (ormid [pointer] == firm){
```

```
    pointer <- pointer + 1};
      {last <- pointer-1};
      }
      if (pointer == numdata+1){last <- pointer-1}
      rel_yvar[first] <- yvar[first];
      for(j in first+1: last){
      rel_yvar[j] = (yvar[j]);
        }
      x <- sort(xtime[first: last])
      y <- length(sort(xtime[first: last]))
      z <- c(1: y)
      y1 <- length(grid)
      h <- dim(as.array(y1))
      for(k in 1: y1){
        for(l in 2: y){
          if(grid[k] > x[l]) {(h[k]<- z[l])}
          if((x[l-1]<= grid[k]) & (grid[k] <= x[l])) {(h[k]<- z[l-1])}
          if(is.na(h[k]))(h[k] <- h[k-1])
          }
        }
      bidfd = smooth.basis(grid, rel_yvar[first-1+h], bidfdPar)
      resultfunc <- bidfd$y;
      vel <- deriv.fd(bidfd$fd,1); #First Deriv. of the function
      acc <-deriv.fd(bidfd$fd,2); #Second Deriv.of the function
###############################################
# Plotting the function, its dynamics, and the raw data

   par(mfrow=c(2,2));
   plot(yvar, xlab="time", ylab="Raw Data");
   plot(bidfd, xlab="time", ylab="Underlying Function");
   par(new=TRUE);
   plot(yvar, xaxt='n', yaxt='n', ann=FALSE);
   plot(vel, xlab="time", ylab="Velocity");
   plot(acc, xlab="time", ylab="Acceleration");

###############################################

# Sample code to perform functional principle component.
pc.y <- princomp(result_function_y, scores = TRUE);

# Sample code to perform functional regression.
lm.y <- lm(result_function_y[t] ~ covariates[t])

# Sample code to perform functional clustering.
y.dist<-dist(result_function_y);
clus.y <-kmeans(y.dist, iter.max=20);
```

FDA libraries in R (http://cran.r-project.org/web/packages/fda/index.html [last accessed on April 20, 2012]) and MATLAB (http://www.bscb.cornell.edu/~hooker/FDAWorkshop/Mintro.m [last accessed on April 20, 2012]) are downloadable.

References

Akdeniz, L., Altay-Salih, A., & Caner, M. (2003). Time-varying betas help in asset pricing: The threshold CAPM. *Studies in Nonlinear Dynamics and Econometrics*, *6*, 1-16.

Bapna, R., Jank, W., & Shmueli, G. (2008). Price formation and its dynamics in online auctions. *Decision Support Systems*, *44*(3), 641-656.

Bertrand, M., & Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, *118*(4), 1169-1208.

Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods*, *7*(4), 400-417.

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*(4), 362.

Bou, J. C., & Satorra, A. (2010). A multigroup structural equation approach: A demonstration by testing variation of firm profitability across EU samples. *Organizational Research Methods*, *13*(4), 738-766.

Boyd, B. K., Gove, S., & Hitt, M. A. (2005). Construct measurement in strategic management research: Illusion or reality? *Strategic Management Journal*, *26*, 239-257.

Brumback, B. A., & Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association*, *93*, 961-994.

Cai, Z. (2011). Functional coefficient models for economic and financial data. In F. Ferraty & Y. Romain (Eds.), *Oxford handbook of functional data analysis*. Oxford, UK: Oxford University Press.

Carmeli, A., & Tishler, A. (2004). Resources, capabilities, and the performance of industrial firms: A multivariate analysis. *Managerial & Decision Economics*, *25*(6-7), 299-315.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal means and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*, 421-483.

Dalal, R. S., & Hulin, C. L. (2008). Motivation in organizations: Criteria and dynamics. In R. Kanfer, G. Chen, & R. Pritchard (Eds.), *Work motivation: Past, present, and future*. New York, NY: Routledge.

Dalal, R. S., Lam, H., Weiss, H. M., Welch, E. R., & Hulin, C. L. (2009). A within-person approach to work behavior and performance: Concurrent and lagged citizenship-counterproductivity associations, and dynamic relationships with affect and overall job performance. *Academy of Management Journal*, *52*(5), 1051-1066.

Dass, M., Jank, W., & Shmueli, G. (2011). Maximizing bidder surplus in simultaneous online art auctions via dynamic forecasting. *International Journal of Forecasting*, *27*(4), 1259-1270.

Davidian, M., Lin, X., & Wang, J. L. (2004). Introduction to emerging issues in longitudinal and functional data analysis. *Statistica Sinica*, *2004*, 613-614.

Fan, J., & Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, *1*, 179-195.

Ferraty, F., & Romain, Y. (2011). *The Oxford handbook of functional data analysis*. New York, NY: Oxford.

Foutz, N. Z., & Jank, W. (2010). Prerelease demand forecasting for motion pictures: Using functional shape analysis of virtual stock markets. *Marketing Science*, *29*(3), 568-379.

Gonzalez-Manteiga, W., & Vieu, P. (2007). Statistics for functional data. *Computational Statistics & Data Analysis*, *51*, 4788-4792.

Grant, R. M. (1996). Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization Science*, *7*(4), 375-387.

Hawawini, G., Subramanian, V., & Verdin, P. (2003). Is performance driven by industry- or firm-specific factors? A new look at the evidence. *Strategic Management Journal*, *24*, 1-16.

Holcomb, T. R., Combs, J. G., Sirmon, D. G., & Sexton, J. (2010). Modeling levels and time in entrepreneurship research: An illustration with growth strategies and post-IPO performance. *Organizational Research Methods*, *13*(2), 348-389.

Hong, Y., & Lee, T. H. (2003). Inference and forecast of exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics*, *85*, 1048-1062.

Ilies, R., Johnson, M. D., Judge, T. A., & Keeney, J. (2011). A within-individual study of interpersonal conflict as a work stressor: Dispositional and situational moderators. *Journal of Organizational Behavior*, *32*(1), 44-64.

Ilies, R., & Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology*, *90*(3), 453-467.

James, G. M., Hastie, T. J, & Sugar, C. A. (2000). Principal components analysis for sparse functional data. *Biometrika*, *87*, 587-602.

James, G. M., & Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, *98*(462), 397-408.

Jank, W., & Shmueli, G. (2006). Functional data analysis in electronic commerce research. *Statistical Science*, *21*(2), 155-166.

Judge, T. A., Scott, B. A., & Ilies, R. (2006). Hostility, job attitudes, and workplace deviance: Test of a multi-level model. *Journal of Applied Psychology*, *91*(1), 126-138.

Kargin, V., & Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, *99*, 2508-2526.

Keats, B. W. (1988.) The vertical construct validity of business economic performance measures. *The Journal of Applied Behavioral Science, 24*, 151-160.

Lang, L. H. P., & Stulz, R. M. (1994). Tobin's q, corporate diversification, and firm performance. *Journal of Political Economy*, *102*(6), 1248.

Le Mens, G., Hannan, M. T., & Pólos, L. (2011). Founding conditions, learning, and organizational life chances: Age dependence revisited. *Administrative Science Quarterly*, *56*(1), 95-126.

Lynall, M. D., Golden, B. R., & Hillman, A. J. 2003. Board composition from adolescence to maturity: A multi-theoretic view. *Academy of Management Review*, *28*(3), 418-431.

Mani, S., Antia, K. D., & Rindfleisch, A. (2007). Entry mode and equity level: A multilevel examination of foreign direct investment ownership structure. *Strategic Management Journal*, *28*(8), 857-866.

McGahan, A. M. (1999). Competition, strategy, and business performance. *California Management Review*, *41*, 74-101.

McGahan, A. M., & Porter, M. E. (2002). What do we know about variance in accounting profitability? *Management Science*, *48*(7), 834-851.

Miller, D. J. (2004). Firms' technological resources and the performance effects of diversification: A longitudinal study. *Strategic Management Journal*, *25*, 1097.

Misangyi, V. F., Elms, H., Greckhamer, T., & Lepine, J. A. (2006). A new perspective on a fundamental debate: A multilevel approach to industry, corporate, and business unit effects. *Strategic Management Journal*, *27*(6), 571-590.

Montemayor, E. F. (1996). Congruence between pay policy and competitive strategy in high-performance firms. *Journal of Management*, *22*(6), 889-908.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies*, *24*(3), 403.

Podsakoff, P. M., MacKenzie, S. B., Jeong-Yeon, L., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879.

Ramsay, J. O. (1990). Economic and financial data as nonlinear processes. In G. P. Dwyer & R. W. Hafer Jr. (Eds.), *The stock market: Bubbles, volatility, and chaos: Proceedings of the Thirteenth Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis* (pp. 81-134). Norwell, MA: Kluwer.

Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and Matlab*. New York, NY: Springer.

Ramsay, J. O., & Ramsey, J. B. (2002). Functional data analysis of the dynamics of the monthly index of non durable goods production. *Journal of Econometrics*, *107*(1-2), 327-344.

Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies*. New York, NY: Springer-Verlag.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York, NY: Springer-Verlag.

Reddy, S. K., & Dass, M. (2006). Modeling on-line art auction dynamics using functional data analysis. *Statistical Science*, *21*(2), 179-193.

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring organizational performance: Towards methodological best practice. *Journal of Management*, *35*(3), 718-804.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York, NY: Oxford University Press.

Sirmon, D. G., Hitt, M. A., & Ireland, R. D. (2007). Managing firm resources in dynamic environments to create value: Looking inside the black box. *Academy of Management Review*, *32*(1), 273-292.

Street, C. T., & Gallupe, R. (2009). A proposal for operationalizing the pace and scope of organizational change in management studies. *Organizational Research Methods*, *12*(4), 720-737.

Wang, S., Jank, W., & Shmueli, G. (2008). Explaining and forecasting online auction prices and their dynamics using functional data analysis. *Journal of Business and Economic Statistics*, *26*(3), 144-160.

You, J., & Jiang, J. (2007). Inferences for varying-coefficient partially linear models with serially correlated errors. In V. Nair (Ed.), *Advances in statistical modeling and inference*. Singapore: World Scientific.

Zhu, H., Brown, P. J., & Morris, J. S. (2011). Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association*, *106*(495), 1167-1179.

Bios

Mayukh Dass is an assistant professor of marketing at the Rawls College of Business, Texas Tech University. He holds a PhD in Business Administration, an MS in Statistics, and an MS in Artificial Intelligence from the University of Georgia. His research focuses on statistical and analytical methods with applications in dynamic economies and brand management. He is a member of the INFORMS and the American Economic Association.

Christine Shropshire is an assistant professor of management at the Terry College of Business, University of Georgia. She received her PhD from Arizona State University in 2008. Her research is in the area of corporate governance, including interlocking directors, the influence of director identities and shareholder preferences on board composition and effectiveness, diversity in the upper echelons, and stakeholder management.