

December 2006

Time Changes Everything: An Examination and Application of Time-Varying Coefficients in Information Systems Research

Eric Overby
Emory University

Benn Konsynski
Emory University

Follow this and additional works at: <http://aisel.aisnet.org/icis2006>

Recommended Citation

Overby, Eric and Konsynski, Benn, "Time Changes Everything: An Examination and Application of Time-Varying Coefficients in Information Systems Research" (2006). *ICIS 2006 Proceedings*. 69.
<http://aisel.aisnet.org/icis2006/69>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2006 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

TIME CHANGES EVERYTHING: AN EXAMINATION AND APPLICATION OF TIME-VARYING COEFFICIENTS IN INFORMATION SYSTEMS RESEARCH

Quantitative Research Methods

Eric Overby

Goizueta Business School, Emory
University

eric_overby@bus.emory.edu

Benn Konsynski

Goizueta Business School, Emory
University

benn_konsynski@bus.emory.edu

Abstract

The Information Systems research field is inherently dynamic. New technologies, new standards, new legislation, and changing user expectations are some of the reasons why topics of interest to the IS field remain in flux. As researchers, we seek to uncover and explain relationships among variables, but due to the dynamism of IS phenomena, these relationships are apt to change over time. For example, the effect of informational features such as product diagnosticity or seller reputation on the price of an electronic commerce transaction is likely to change over time as users become more comfortable with online trading. This paper describes several statistical methods to model these changes in relationships. Specifically, we discuss methods to investigate time-varying coefficients in regression models, including rolling regression, "parameterizing" the coefficients using process functions, and testing for structural change. Importantly, we describe how the structure of many of the data sets used in IS research differs from that of data sets often used in other fields such as finance, economics, or marketing. This has implications for the investigation of time-based effects. We illustrate each method using a data set gathered from the wholesale automotive market, which not only helps us explain the methods, but also allows us to investigate the evolution of market practice in one empirical context. Thus, we address both methodological and substantive issues. Given that our field is inherently dynamic, an understanding of how effects change over time should be central to the overall IS research agenda. This paper is designed to familiarize IS researchers with methods available for this purpose.

Keywords: time-varying coefficients, time-varying parameters, varying parameter regression, Chow test, CUSUM/MOSUM, rolling regression, moving regression, process function, electronic commerce, electronic markets, automotive industry.

Introduction

If there is one constant in the Information Systems field, it is that things rarely remain constant. Much of what we study is in flux. Changes occur as new information technologies are introduced, existing technologies mature, society grows more accepting (and demanding) of information technology, and new technologically-enabled business models are launched and proven. Due to the rate of change in the phenomena of interest to the IS field, it is likely that relationships among variables of interest to IS researchers change over time. For example, the influence of variables such as product diagnosticity or seller trustworthiness on the price of an electronic commerce transaction may change over time as individuals and firms become more experienced with online trading (Ba and Pavlou 2002; Dellarocas 2003; Jiang and Benbasat 2004; Kauffman and Wood 2005). Similarly, the influence of an information system on an element of firm performance, such as customer satisfaction, may change over time as the firm assimilates the information system into its work practices (Fichman and Kemerer 1997). Although there are

numerous opportunities to investigate how relationships among variables evolve over time, these opportunities are too often left to lie fallow, or else shunted into the purgatory we refer to as “future research opportunities.”

The purpose of this paper is to review several available statistical methods for determining if relationships among variables change over time. In other words, we describe methods of testing for time-varying coefficients. Our treatment of this subject is practically oriented and not technical in nature. We believe this paper will be useful for IS researchers who seek to investigate the exciting dynamics inherent to the IS field. The purpose of the paper is not to critique existing research practice but rather to influence future practice. It is not that IS researchers are misapplying techniques to investigate relationships over time; it is that they are too often not applying them at all! This may be a function of data limitations, but it may also be due to difficulties in accessing the appropriate methodology. Some of the most common methods for investigating time-based effects are not appropriate for many types of IS data, which may frustrate IS researchers and cause them to set aside analysis of time-based effects. This paper seeks to remedy this situation. We describe why the structure of data sets often found in IS research differs from the structure of data sets typically encountered in other fields such as finance or marketing, and how this influences the types of statistical methods that are appropriate for many IS research problems.

In addition to discussing statistical methods well-suited for the structure of many IS data sets, we illustrate their use through the analysis of electronic markets data gathered from the wholesale automotive industry. This data is ideal for illuminating the methods discussed herein, because the influence of several variables on electronic market practice is likely to change as the electronic marketplace continues to evolve. The conclusions we draw from the data analysis make this not only a methodological paper, but also a substantive one, as we are able to make statements about the evolving impact of information technology in one empirical setting.

We begin by delineating what this paper is, and is not, about, with a particular focus on how several IS data sets are often structured differently than data sets found in other fields. The peculiarities of several IS data sets render many of the most common methods for investigating time-based effects inappropriate. After describing the wholesale automotive data that provides our empirical context, we describe several statistical methods that are appropriate for testing for time-varying coefficients in IS research. Methods discussed include tests for structural change such as the Chow test and the CUSUM/MOSUM test; rolling regression; and “parameterizing” the coefficients using process functions. The empirical context allows us to illustrate each method and highlight its strengths and weaknesses, as well as make substantive statements about the evolution of market practice in the wholesale automotive industry. We close with a discussion of how this line of discussion might, itself, evolve over time.

The Peculiarities of Several IS Data Sets: Pooled Cross-Sections versus Panel or Time Series Data

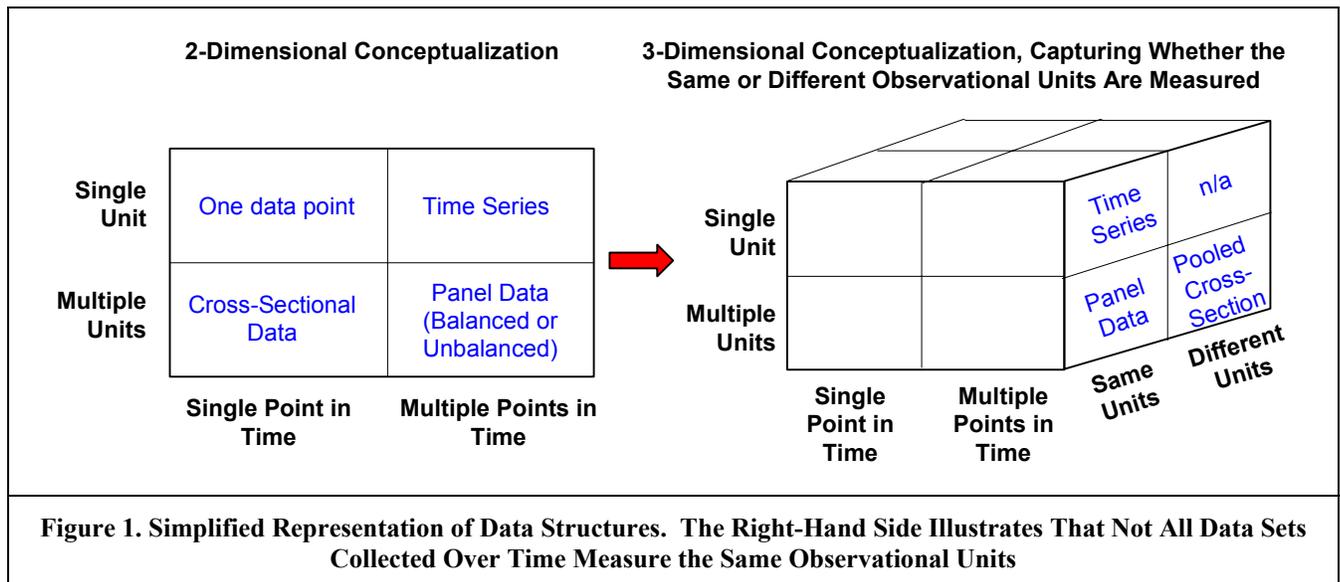
Consider the type of data that is often available and used in IS research. For example, IS researchers were among the first to collect large electronic commerce transaction data sets by downloading publicly available data from the Internet using software agents. Combining technological know-how with research curiosity, IS researchers amassed data sets consisting of, for example, eBay transactions, Amazon pricing and sales data, airline prices, hotel prices, etc. (e.g. Ghose et al. 2006; Kauffman and Wood 2000). IS researchers have obtained access to clickstream data (Chen and Hitt 2002) and transaction data that isn’t publicly available (Banker and Mitra 2005; Lee 1998). Some have analyzed the usage logs of electronic mail systems, ERP systems, and other types of enterprise systems (Devaraj and Kohli 2003).

In each of these types of data, it is common to have observations that occur at different points in time. For example, in an eBay transaction data set scraped from the Internet, some of the transactions may have occurred in January and others in September. Similarly, the user sessions reflected in the usage logs of enterprise systems will have occurred at different times. Thus, it is possible to investigate how relationships within these data sets evolve over time.

Many of the commonly used statistical methods for investigating relationships over time are designed for panel or time series data. The difficulty facing IS researchers is that many data sets used in IS research are not structured as panels or time series, rendering these methods inappropriate. We will digress briefly to distinguish between panel data, time series data, and the type of data of interest in this paper, which we refer to as pooled cross-sectional data. A *panel data set* is one that consists of repeated observations of the same units over time (Wooldridge 2002). For example, the annual IT spending between 2000 and 2005 for 50 firms would constitute a panel data set. If data for each year is available for all 50 firms, the panel is called a *balanced panel*. If some of the firms are missing

observations in certain years, the panel is called an *unbalanced panel*. Some researchers use the term *longitudinal* data synonymously with panel data (Wooldridge 2002). A *time series* is defined as a collection of observations made sequentially in time (Chatfield 1996). For example, the number of visitors to a corporate Web site measured on a weekly basis represents a time series. If only a single variable is measured, the time series is referred to as *univariate*. If multiple variables are measured, such as the number of Web site visitors and the number of page views, the time series is referred to as *multivariate*. A key assumption about time series is that preceding observations are correlated with subsequent observations. In other words, the observations are not independent, and the order of the observations is meaningful. For example, a good predictor of how many people will visit a Web site this week is the number of people who visited the Web site last week. Another characteristic of time series is that observations are usually recorded at regular intervals, such as daily, weekly, monthly, or annually. Note that multiple time series measured over the same periods results in the same structure as panel data. For example, if a time series measuring the number of Web site visitors on a weekly basis was collected for 50 firms, the combined data would have a panel structure. Panel and time series data are common in fields such as economics, accounting, finance, and marketing. For example, accounting and finance researchers often analyze stock price data that has been gathered on a fixed group of stocks on a daily basis. Marketing researchers often examine brand sales data in which the sales of a fixed group of brands are measured weekly or monthly.

The left-hand side of Figure 1 provides a simplified representation of the relationship between panel, time series, and *cross-sectional* data, which is data collected at a given point in time (Wooldridge 2002). Cross-sectional data consists of observations of multiple units at a single point in time. A time series consists of a single unit measured at multiple times, and a panel consists of multiple units measured at multiple times. (If multiple time series are collected at the same intervals, the distinction between time series and panel data blurs, as mentioned above.) Although this schematic is intuitively appealing, it omits an important dimension. The right-hand side of Figure 1 incorporates a third dimension into the schematic to capture whether the same or different units are observed over time. An assumption for both panel and time series data is that the same units are measured over time, for example, the same 50 firms mentioned in the examples above. This assumption does not hold in many of the data sets of interest to IS researchers, as these data sets often consist of observations of *different* units over time. The term *pooled cross-sectional data* refers to data consisting of different observational units at different points in time (Wooldridge 2002). For example, units A and B may be observed in the cross-section taken at time 1, units C and D may be observed in the time 2 cross-section, etc. If the same units are observed at multiple points in time, this is taken to be coincidental, rather than by design as in panel or time series data.



Several data sets of interest to IS researchers are pooled cross-sections. In other words, they consist of observations across time, but these observations are of different units. First, consider electronic commerce transaction data, such as that scraped from eBay. IS researchers often focus on the individual user as the unit of analysis in these data sets, in order to analyze their behavior. For example, an IS researcher might be interested in how informational features

such as seller reputation score or product diagnosticity affect a user's willingness to pay. Because it is typically uncommon for the same user to appear multiple times in the same data set, the data is not structured as a panel. Instead of consisting of the same unit measured at multiple times, they consist of *different* units measured at multiple times. Second, consider clickstream data, which tends to be structured similarly to electronic commerce transaction data in that it contains observations of different units over time. In addition, clickstream data is often de-identified, so that even if the same unit appears multiple times in the data, it may be impossible to distinguish that unit from the others. Cookies and user logins may help mitigate this issue, but they are indeterminate, as it is relatively easy (and often desirable) for Internet users to browse anonymously (for example, by disallowing cookies.) Similar issues may arise in the analysis of other types of usage logs, such as those for enterprise systems. Third, consider survey data, such as that measuring firm or individual IT usage, collected at multiple points in time. It may be infeasible for the researcher to survey the same respondents at each point in time. Thus, it is often the case that IS data cannot be modeled as a classical time series or as a panel data set, because it is not possible to track the same unit over time. (Some would argue that the types of IS data sets mentioned above could be classified as unbalanced panels, but using the term panel to refer to the IS data sets of interest in this paper is misleading, as they tend to lack any underlying panel structure.)

Whether or not a data set is structured as a pooled cross-section, a panel, or a time series has important implications for the appropriate statistical methodology. Methods designed for the analysis of time series and panel data are based on the assumption that the same units are measured over time. This means that the observations for a given unit in a data set are not independent. For example, in a panel data set consisting of IT spending for 50 firms over 6 years, the observations for each firm are assumed to be correlated, as it is likely that there are firm-specific factors such as industry membership and management skill that influence a given firm's IT spending in a similar manner each year. Fixed effects or random effects are often used to capture these unit-specific influences, so that the other coefficients of interest can be estimated separately. (In short, *fixed effects* estimation assumes that the unit-specific effect is correlated with the other explanatory variables, while *random effects* estimation assumes that the unit-specific effect is uncorrelated with the other explanatory variables. Readers interested in a more nuanced description of the difference are referred to Wooldridge (2002: p. 251).) In a time series, successive observations are typically assumed to be correlated (Chatfield 1996). This allows a time series to display a trend. A variety of filters are used to differentiate the systematic from the stochastic elements of the trend, including autoregressive ("AR") filters and moving average ("MA") filters. (Briefly, an *autoregressive filter* relates current observations to previous observations, while a *moving average filter* relates current observations to previous shocks. AR and MA processes are often used in combination, yielding autoregressive moving average ("ARMA") and autoregressive integrated moving average ("ARIMA") models. See Enders (2004) and Chatfield (1996) for additional description of these methods.)

Because these methods assume that observations are correlated, they are typically not appropriate for the types of pooled cross-sectional data often of interest to IS researchers. This is because there is no a priori expectation in pooled cross-sectional data that the observations are correlated, which is defensible given that the observations are of different units. Thus, an IS researcher attempting to apply time series or panel methods to a pooled cross-sectional data set may quickly become frustrated. In the remainder of the paper, we focus on a set of statistical methods that are appropriate for analyzing time-varying relationships within the types of pooled cross-sectional data often of interest to IS researchers. These methods can be used to detect time-based change in an entire set of coefficients as well as in a specific coefficient. We limit our inquiry to regression models. We hope that this discussion will complement the more widely available discussions on investigating time-varying relationships in time series and panel data (e.g. Enders 2004; Hamilton 1994).

Modeling Time-Varying Coefficients in Pooled Cross-Sectional Data

The Empirical Context Used to Illustrate the Methods

Before embarking on our discussion of the methods available to investigate time-varying coefficients in pooled cross-sectional data, we first introduce the empirical context we use to illustrate the methods. The empirical context is the used automobile wholesale market. This market facilitates the exchange of used vehicles between institutional sellers and buyers. Sellers include rental car companies, the financial affiliates of automotive manufacturers, and other fleet operators who wish to sell large quantities of used vehicles in the wholesale market. Buyers are typically

licensed automobile dealers who seek to purchase used vehicles in the wholesale market for resale to the consumer public. For example, in order to dispose of vehicles no longer appropriate for rental, a rental car firm may use the wholesale automotive market to sell hundreds of late model vehicles from its fleet. Automobile dealers use the market to purchase the vehicles at wholesale prices, which they then resell to the consumer public at retail prices. There are several intermediaries in this market that provide a range of market-making services, such as arranging for an ample supply of vehicles, attracting a mass of potential buyers, certifying vehicle quality, facilitating price discovery, and providing transaction support.

As this is a wholesale market, the consumer public is rarely allowed to participate. Like many business-to-business markets, the used automotive wholesale market operates “beneath the radar,” although the number of transactions and dollar value of the vehicles exchanged are staggering. In 2005, approximately 23 million used vehicles were exchanged in this market, for a total of approximately \$195 billion.¹ (Figures are for the United States.) Thus, the happenings within this market are not only interesting from a research perspective, but also have a sizable impact on the overall economy.

The wholesale automotive market has traditionally operated as a physical market in which buyers, sellers, and vehicles are all collocated at a physical facility where sales events are held. In a typical sales event, hundreds of vehicles are driven, one at a time, into the midst of a group of buyers, who then bid on and purchase the vehicles via an ascending auction process. Recently, many of the players in this market, most notably the sellers and the intermediaries, have introduced new electronic mechanisms to the market. For example, vehicles no longer have to be physically collocated at the market facility to be offered for sale; they may instead be offered electronically. If a vehicle is presented physically, it is driven through the market facility as described above. If a vehicle is presented electronically, a photograph of it, along with textual information, is displayed on a screen in the market facility. An unusual feature of this market is that both the physical and electronic vehicle presentation mechanisms are used in the same sales events. For example, if a sales event consists of 200 vehicles, the first five might be presented physically, the next five electronically, and so on.

The data set consists of 11,208 transactions between buyers and sellers in this market. These transactions occurred in 79 discrete sales events between August 2004 and January 2005. All sales events were facilitated by the same intermediary. In approximately 17% of these transactions, the vehicle was presented electronically. (In the other 83%, the vehicle was presented physically.) We have analyzed the relationship between vehicle presentation method (physical vs. electronic) and the price that buyers pay. We refer to this relationship as the coefficient for the ELECTRONICVEHICLE variable. We have coded ELECTRONICVEHICLE as a dummy variable. ELECTRONICVEHICLE is set to 1 for vehicles presented electronically and 0 for vehicles presented physically. The ELECTRONICVEHICLE coefficient is likely to vary over time as market participants become accustomed to purchasing electronically presented vehicles. For example, because buyers have traditionally purchased physically presented vehicles, they may be initially wary of purchasing electronically presented vehicles. This could cause them to discount what they pay. However, as buyers become more comfortable with this presentation method, they may cease to apply discounts. In that case, the ELECTRONICVEHICLE coefficient might start out as a negative number, but then converge toward 0 over time. We investigate this and related hypotheses about this coefficient using each of the methods for testing for time-varying coefficients described below. In order to isolate the effect of ELECTRONICVEHICLE on price, we have collected multiple control variables related to vehicle and sales event characteristics. For example, we control for each vehicle’s condition (measured on a 0-5 scale by the intermediary, where 0 represents a vehicle suitable only for scrap parts and 5 a near-pristine vehicle) and wholesale valuation (as established by the intermediary based on a vehicle’s year, make, model, and mileage). We also control for sales event characteristics such as the number of buyers who participate. These controls allow us to separate the effect attributable to vehicle presentation from other confounding variables. Additional detail about this data set, the empirical context, model development and testing (including summary statistics for dependent, independent, and control variables and additional detail on the econometric techniques employed), theoretical linkages, etc., is available in Overby and Jap (2006). Much of this detail is withheld from this manuscript due to space limitations and this paper’s focus on testing for time-varying coefficients.

This data represents pooled cross-sectional data. The cross-sections are taken at multiple points in time between August 2004 and January 2005. The unit of analysis is the vehicle (identified by its Vehicle Identification Number

¹ Figures are from the National Auto Auction Association (www.naaa.com.)

("VIN".)) Each VIN is observed in the data set only once. Thus, the data set consists of observations of different units over time.

Note that it would be possible to convert this data into a panel if we changed the unit of analysis from individual vehicles as identified by their VINs to vehicle groups based on make/model combinations (e.g. Ford Taurus, Honda Accord, etc.) This is because although we do not observe the same VIN's in the data set over time, we do observe the same make/model combinations over time. Table 1 provides an illustration of how this data set could be "panelized." The main drawback to conducting the analysis in this manner is that it throws away data by assuming that all vehicles of a given make/model are homogeneous. The advantage is that it opens up additional analysis possibilities, as panel data methods become viable. Make/model combination represents a natural way to "panelize" this data set, but in the absence of such a logical choice, propensity scoring methods may be used to group observations into classes based on their attributes.

Table 1. Two Ways to Structure the Wholesale Automotive Data

Native Format (Pooled Cross-Sectional)				Transformed Format (Restructured as a Panel)			
Unit of analysis is the vehicle as specified by VIN. Each VIN is observed only once, at a specific time period (e.g. T1, T2, or T3.)				Unit of analysis is the make/model combination. Each make/model combination is observed multiple times, in multiple time periods.			
	T1	T2	T3		T1	T2	T3
1FAFP52U9WA165605 (Ford Taurus)	✓			Ford Taurus	✓	✓	✓
1FAHP56S02A121156 (Ford Taurus)			✓	Honda Accord	✓	✓	✓
1FAHP56SX4A219705 (Ford Taurus)		✓					
1HGCG56611A112869 (Honda Accord)		✓					
JHMCG5665YC038981 (Honda Accord)	✓						
1HGCG5641YA127413 (Honda Accord)			✓				

The Methods

We present several methods to test for time-varying coefficients. These methods apply to testing both continuous and discrete changes in coefficients, and can be used to test changes in entire coefficient sets or in a specific coefficient. The methods we discuss are:

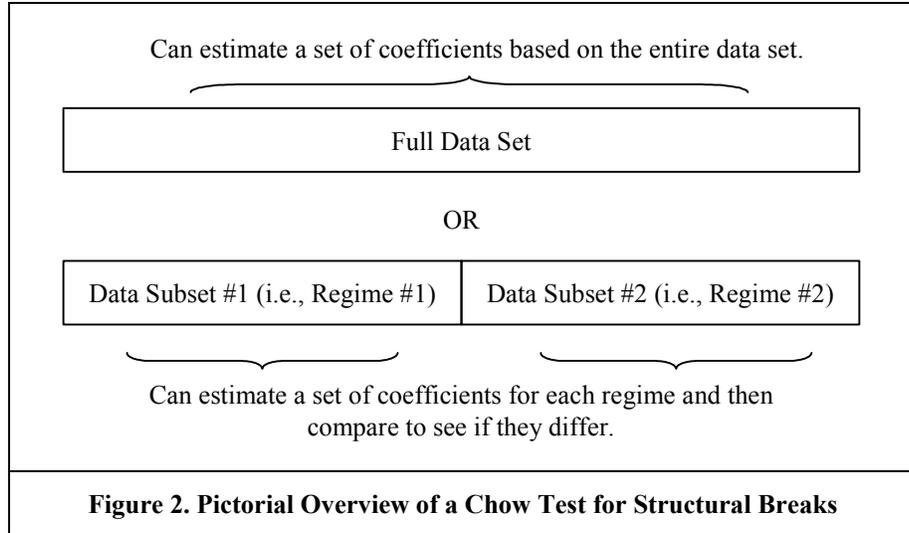
- Tests for structural change, including the Chow test and the CUSUM/MOSUM test,
- Rolling regression, and
- “Parameterizing” the coefficients using process functions.

The tests for structural change and rolling regression are relatively discrete, stand-alone, methods. By contrast, parameterizing the coefficients represents an entire class of methods. Several methods for examining time-varying coefficients can be classified as specific instantiations of the general idea of parameterizing the coefficients. Thus, structuring the discussion in this manner allows us to discuss a myriad of methods in a concise and accessible way.

Tests for Structural Change

Tests for structural change are useful to diagnose whether coefficients are stable within a data set or if they shift, either as a function of time or some other variable. Two common methods to test for structural change are the Chow test and the CUSUM/MOSUM test. (CUSUM is short for “Cumulative Sum” and MOSUM for “Moving Sum.”) In the context of time-varying coefficients, the Chow test is useful if the coefficients are believed to have shifted after some specific event, while the CUSUM/MOSUM test is useful to diagnose when a shift might have occurred.

The Chow Test: The Chow Test is commonly used to test for “structural breaks” within a data set (Chow 1960). It can be used to compare the coefficients associated with a subset of a data set to those associated with a different subset. The Chow test can be useful to test for time-varying coefficients, as it allows a researcher to test whether coefficients at one point in time are different from those at another point in time. Figure 2 provides a pictorial representation of how a Chow test can be used to estimate coefficients from different subsets within a data set (referred to as “regimes”), which can then be compared to one another.



The first step in conducting a Chow test is to identify the event after which the coefficients are hypothesized to have changed. Examples of such events in IS research include the upgrade of an enterprise system or the introduction of a new Web site feature. The coefficient describing the relationship between, for example, usage intentions and actual usage for these information systems might differ before and after the event. Once the event has been identified, the next step is to divide the data into two sub-samples, referred to as *regimes*. The first regime contains all observations before the event, and the second regime contains all observations after the event.

The Chow test statistic is given by $\frac{SSR_{pooled} - (SSR_1 + SSR_2) / k}{SSR_1 + SSR_2 / (n - 2k)}$, where SSR_{pooled} is the sum of the squared

residuals from the regression using the entire sample, SSR_1 and SSR_2 are the sum of squared residuals from regressions using each individual regime, k is the number of variables in the regression model, and n is the total number of observations in the overall sample. The SSR terms can be computed by squaring and summing the residuals generated by each regression. After calculating the Chow test statistic, the next step is to compare it to a critical value to assess significance. The critical value is drawn from the F-distribution with k and $n-2k$ degrees of freedom at the appropriate significance level (usually 95%). If the Chow test statistic exceeds the critical value, then the coefficients are believed to vary between the two regimes.

A drawback to the Chow test is that it is sensitive to how the regimes are delineated. In some cases, a specific event may suggest a natural breakpoint. However, in many cases, it is difficult to determine, a priori, what constitutes the appropriate breakpoint. For example, there is no event that represents a natural breakpoint in our empirical context. However, we can still use the Chow test with our data for diagnostic purposes. To illustrate, we first divide our data set into two regimes. The “Early” regime consists of the observations during the first three months of the data set, and the “Late” regime consists of the observations during the last three months of the data set. A Chow test indicates that the coefficients in the Late regime are significantly different than those in the Early regime. The Chow test statistic = 2.56, which is significant at the 1% level. This provides some evidence that the coefficients are evolving, but doesn’t provide much insight into the process of the evolution, or into which specific coefficients might be evolving. These questions can be analyzed via other methods presented in this paper.

The CUSUM/MOSUM Test: In contrast to the Chow test, the CUSUM/MOSUM test does not require a priori specification of a breakpoint. We first discuss the CUSUM test, and then distinguish CUSUM from MOSUM.

The CUSUM test uses the residuals from a regression to determine if coefficients are stable or varying over time. Assume that we have n observations, ordered by when they occurred. First, we fit a regression model to the first k observations, $k < n$. We use the coefficient estimates from the first k observations to predict the value of observation $k + 1$. We then measure the accuracy of the prediction by examining the residual, which is the difference between the predicted value and the actual value. This process is repeated as k grows. If the coefficients

are stable, then each prediction will be approximately as accurate as the previous prediction. However, if the coefficients change over time, then predictive accuracy will decrease, because the coefficients that provide good predictions in the first part of the data will not provide good predictions in the latter part of the data. This is reflected in an increase in the absolute value of the residuals.

Figure 3 shows a plot of the normalized residuals obtained from a CUSUM procedure on our data set. Predictions appear relatively stable in the first 25% of the observations, after which predictive accuracy begins to drop. Predictions stabilize again towards the end of the data set. This provides additional evidence that the coefficients vary over time.

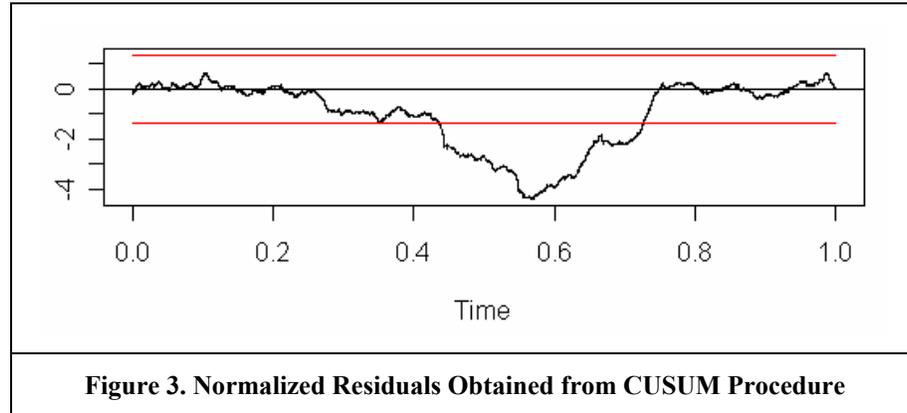


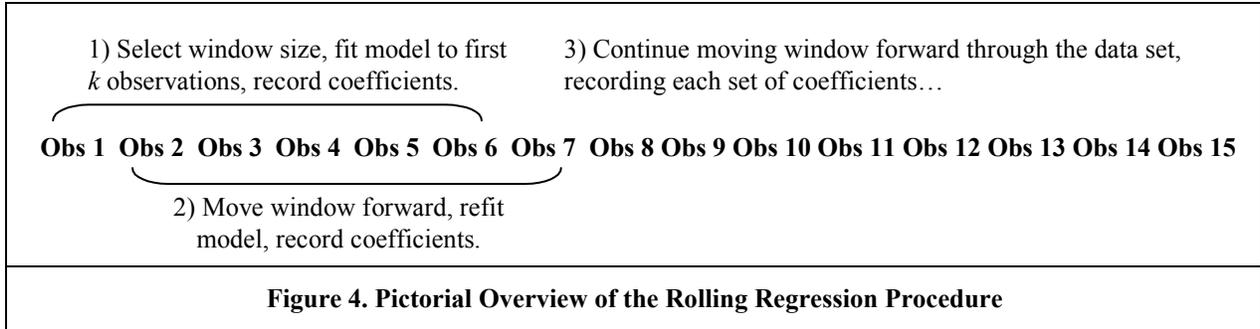
Figure 3. Normalized Residuals Obtained from CUSUM Procedure

The CUSUM procedure uses all previous observations to predict the next observation. By contrast, the MOSUM procedure uses only the previous g observations. In other words, MOSUM predicts each observation based on a rolling window of prior observations. The original CUSUM procedure is attributed to Brown, Durbin, and Evans (1975). A more recent discussion of models to detect structural change, including CUSUM/MOSUM procedures and the Chow test is provided by Andrews (1993). Both the CUSUM and MOSUM procedures are available in statistical packages. For example, they are both implemented in the *strucchange* package for the R program (<http://www.r-project.org/>.)

These types of models are useful to detect whether entire sets of coefficients change over time, but they are not helpful if a researcher is interested in whether a specific coefficient changes over time, unless the regression model only contains a single coefficient. We now turn the discussion to models designed to detect changes over time in a specific coefficient.

Rolling Regression

Rolling regression, aka moving regression or moving window regression, is one method to investigate time-based effects in a specific coefficient. Rolling regression fits a regression model to a data set multiple times by moving forward through the data set in a rolling fashion (Brown et al. 1975). For example, assume a data set consists of 1,500 observations at different points in time. In a rolling regression procedure, the researcher first orders the observations from earliest to latest, and then fit a regression model to the first k observations. Let $k = 100$. This means that the first regression is fitted using observations 1 through 100. The researcher records this initial set of coefficient estimates and then fits the same regression model using observations 2 through 101. After recording this set of coefficient estimates, the researcher fits the model using observations 3 through 102, and so on until the end of the data set. The parameter k is referred to as the *window size*. Rolling regression produces multiple sets of coefficient estimates, depending on the window size, the number of observations in the data set, and the *step size*, which is the increment by which the window is moved each iteration. Assuming the researcher rolls forward a single observation at a time (i.e., the step size = 1), the number of sets of coefficient estimates produced is $n - k + 1$, where n represents the number of observations. Rolling regression procedures are available in many statistical packages. For example, in STATA the procedure is entitled “rolling.” Figure 4 provides a pictorial overview of the rolling regression procedure.



We illustrate a rolling regression using the wholesale automotive data by first ordering the 11,208 transactions according to which occurred first. We then select the window size. Selecting the window size is something of an art, as the window should be large enough for each regression to have adequate power but small enough to reflect changes in the coefficients (otherwise the estimates will too closely resemble the estimates for the sample as a whole.) Thus, performing sensitivity analyses based on different window sizes is advisable. We use a window size of 3,000, which generates 8,209 sets of coefficient estimates ($11,208 - 3,000 + 1 = 8,209$.) (We also used window sizes of 1,000 and 5,000 and received similar results.) By plotting the coefficient estimates against time, we can get a sense of how the coefficient may be changing. Figure 5 displays a plot of the ELECTRONICVEHICLE coefficient.

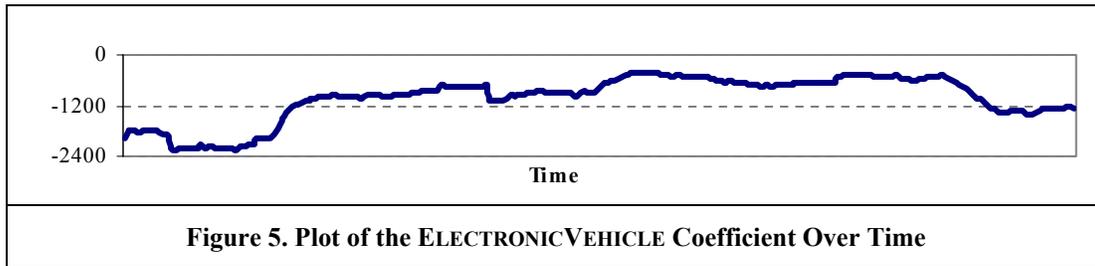


Figure 5 suggests that the ELECTRONICVEHICLE coefficient has diminished in absolute value over time. At the beginning of the time span, buyers discounted electronically presented vehicles by approximately \$2,000. Over the balance of the time span, this discount was closer to \$800, generally ranging between \$400 and \$1,200. This suggests that buyers may have initially been skeptical of electronic vehicles and heavily discounted what they paid for them. However, as buyers gained experience with this new vehicle presentation method, the discounts diminished.

Rolling regression overcomes one of the key limitations of the Chow test, in that it does not require the researcher to specify, a priori, any breakpoints or regimes. Rolling regression permits a view of how coefficients evolve over time on a continuous basis. Rolling regression is particularly useful if the observations are spread relatively evenly through time. In other words, the continuous nature of the rolling regression procedure is best leveraged when used with a data set whose observations are spread more or less continuously through time. If, on the other hand, most of the observations are “clumped” at specific periods in time, rolling regression may suggest a continuity in a coefficient’s evolution that doesn’t exist, especially if there are only a few clumps.

A disadvantage to rolling regression is that it requires a relatively large data set, as the number of observations in the data set should be several times larger than the window size to capture the dynamism in a coefficient. This is because if the window size is too large relative to the overall number of observations, then coefficient estimates from each window will too closely resemble those obtained from the data set as a whole. However, the window size should be large enough to provide adequate statistical power, so that the coefficient estimates for each iteration are stable. Another disadvantage is that rolling regression does not provide formal statistical tests beyond whether or not the coefficient for a given window is statistically different from zero. For example, in our empirical context, we cannot formally test whether the estimates from the first part of the data set (which are -2000 on average) are statistically different from the estimates in the balance of the data set (which are -800 on average.)

Rolling regression is one method to investigate evolution in coefficient estimates over time. Another method is to “parameterize” the coefficients via process functions.

Parameterizing the Coefficients Via Process Functions

This section presents methods of parameterizing coefficients to test for both continuous and discrete changes in a coefficient over time. These methods are very general and can be used to capture a myriad of ways a coefficient might evolve.

Modeling a continuous change in a coefficient: Traditional regression models assume that the value of a dependent variable is a function of independent variables, whose influence on the dependent variable is weighted according to regression coefficients. In other words,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (\text{eq. 1})$$

where y is the dependent variable, x_1 to x_k are independent variables, β 's are regression coefficients, and ε is an error term. In many regression models, little attention is paid to the structure of the β 's. They are assumed to be stationary and to lie within some confidence interval (which allows us to test for statistical significance), but they are not considered to be functions of other variables.

But consider the possibility of a time-varying beta coefficient. It is possible to represent this beta coefficient as a function that depends on time (Farley and Hinich 1970; Hinich and Roll 1981). Such a function is often referred to as a *process* function (Gatignon and Hanssens 1987; Naik and Raman 2003; Wildt and Winer 1983). For example, we can model β_1 as being a function of time (denoted as t) in the following manner:

$$\beta_{1,t} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_k t^k + \omega, \text{ where } \omega \text{ is an error term.} \quad (\text{eq. 2})$$

In equation 2, $\beta_{1,t}$ is modeled as a polynomial of degree k . α_0 represents the baseline (time-invariant) component of β_1 , while the other terms capture how β_1 varies with time. The null hypothesis of $\alpha_1 = \alpha_2 = \dots = \alpha_k = 0$ would suggest that the coefficient is unaffected by time and is therefore, time-invariant. However, rejecting this null hypothesis would suggest that the coefficient is a function of time, and the values of the α coefficients would suggest how. For example, if α_1 is significant and positive, then the β_1 coefficient would increase linearly with time. A positive and significant α_2 coefficient would suggest that β_1 is increasing at a growing pace over time. Alternative process functions for β_1 can be constructed based on the researcher's theory of how the coefficient might vary with time. For example, equation 3 could be used if the researcher expects the coefficient to increase with time, but at a declining rate.

$$\beta_{1,t} = \alpha_0 + \alpha_1 \ln(t) + \omega \quad (\text{eq. 3})$$

Assume we have constructed the regression equation shown in equation 1, but believe that the β_1 coefficient varies over time according to the process function shown in equation 2. We limit equation 2 to be a second-degree polynomial for simplicity. We write down equation 1, substituting equation 2 in place of the β_1 term, which yields:

$$y = \beta_0 + [\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \omega] x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (\text{eq. 4})$$

By multiplying x_1 through the expression for β_1 , we get:

$$y = \beta_0 + \alpha_0 x_1 + \alpha_1 t x_1 + \alpha_2 t^2 x_1 + \omega x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (\text{eq. 5})$$

Effectively, we have introduced interaction terms into our regression and complicated the error structure by introducing ω . If we assume that the process function is undisturbed by an error term, in other words, if we remove ω , equation 5 reduces to:

$$y = \beta_0 + \alpha_0 x_1 + \alpha_1 t x_1 + \alpha_2 t^2 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (\text{eq. 6})$$

We can estimate equation 6 in a straightforward fashion. α_0 is the coefficient for the x_1 variable, α_1 is the coefficient for the interaction between the x_1 variable and the time variable, and α_2 is the coefficient for the interaction between

the x_1 variable and the time variable squared. Inspection of the α_1 and α_2 variables can indicate whether the β_1 coefficient varies over time in accordance with the process function we have specified.

To illustrate this method in our empirical context, we first construct a process function for the ELECTRONICVEHICLE coefficient, which we refer to as β_1 . Suppose we believe the process function for β_1 to be best represented as a logarithmic function such that β_1 increases over time but at a declining rate. Let:

$$\beta_{1,t} = \alpha_0 + \alpha_1 \ln(t) \tag{eq. 7}$$

As described above, this introduces an interaction term into our model: the interaction between the ELECTRONICVEHICLE variable and the natural log of the TIME variable. (TIME ranges from 1 to 174 and represents which day within the 6-month time span a transaction occurred.) Thus, after substituting eq. 7 in place of β_1 , our model becomes:

$$y = \beta_0 + \alpha_0 x_1 + \alpha_1 \ln(t) * x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \tag{eq. 8}$$

Table 2 lists the coefficients for this regression model.

Table 2. Coefficients Estimates Assuming a Continuous Change in the ELECTRONICVEHICLE Coefficient			
	Coefficient Estimate	Standard Error	p-Value
β_0 : INTERCEPT	-2209.14	282.28	0.00
α_0 : ELECTRONICVEHICLE	-2433.37	726.84	0.00
α_1 : ELECTRONICVEHICLE * LN(TIME)	317.82	177.50	0.08
β_2 : VEHICLECONDITION	725.97	69.32	0.00
β_3 : VEHICLEVALUATION	0.95	0.01	0.00
β_4 : NUMBERBUYERS	3.74	2.14	0.09
$R^2 = 0.98$ (Note that the VALUATION variable is highly predictive, as the intermediary is known for its precision in estimating wholesale vehicle values.)			

Table 2 indicates that the ELECTRONICVEHICLE * LN(TIME) coefficient, which is our estimate of α_1 , is positive and significant. This provides additional evidence that the coefficient is increasing over time. This may be because buyers became more comfortable with the electronic vehicle presentation mechanism as they gained direct experience or as positive word-of-mouth spread, causing them to discount less.

An advantage to the parameterizing the coefficients method is that it allows the researcher to explicitly model how the coefficient varies with time. This is a step beyond the rolling regression method, which can reveal whether the coefficient evolved over time, but provides little insight into the pattern of evolution. Parameterizing the coefficients also permits formal tests of whether the coefficient is changing over time. For example, it is straightforward to test whether process function coefficients are statistically significant. In the model above, this can be done via a t-test to determine whether α_i is statistically significant. A disadvantage of parameterizing the coefficients is that the researcher must choose the appropriate process function for the coefficient. It is not always clear which function is best, although theory and model fitting techniques can provide guidance. Another drawback is the number of interaction terms introduced to the model, particularly if the researcher is interested in testing whether more than one coefficient varies over time. This can lead to problems with multicollinearity and statistical power. Last, assuming that the process function can be defined without an error term is an assumption that, while useful for model implementation and interpretation, may be unfounded. If the process function does include an error term, GLS methods may be used to estimate the model, as they can account for the increased complexity of the error term.

Modeling a discrete change in a coefficient: The above description applies if a given coefficient is believed to evolve continuously over time. Other methods are useful if the coefficient is thought to have changed as a result of a discrete event, such as an enterprise system upgrade or the passage of new Internet-related legislation. These

methods are closely related to the Chow test described above, as they involve examining whether a coefficient differs across regimes.

To illustrate using our empirical context, we first divide the data set into regimes. The “Early” regime consists of the observations during the first three months of the data set, and the “Late” regime consists of the observations during the last three months of the data set. Let z represent a dummy variable which is equal to 0 for observations in the Early regime and 1 for observations in the Late regime.

Assume that our base regression model is:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (\text{eq. 9})$$

We define the following process functions for β_0 and β_1 .

$$\beta_0 = \alpha_0 + \alpha_1 z \quad (\text{eq. 10})$$

$$\beta_1 = \alpha_2 + \alpha_3 z \quad (\text{eq. 11})$$

Substituting equations 10 and 11 into equation 9 yields:

$$y = \alpha_0 + \alpha_1 z + \alpha_2 x_1 + \alpha_3 (z * x_1) + \dots + \beta_k x_k + \varepsilon \quad (\text{eq. 12})$$

For observations in the Early regime, $\alpha_1 z$ and $\alpha_3 (z * x_1)$ will drop out, as $z=0$ for these observations. This means that the coefficient for the x_1 variable in the Early regime is simply α_2 . However, for observations in the Late regime (where $z=1$), the $\alpha_1 z$ and $\alpha_3 (z * x_1)$ terms reduce to α_1 and $\alpha_3 x_1$. This leaves two terms involving x_1 : $\alpha_2 x_1$ and $\alpha_3 x_1$. By collecting terms, we can see that the coefficient for the x_1 variable in the Late regime is $\alpha_2 + \alpha_3$. Whether the coefficient in the Early regime is different from the coefficient in the Late regime, in other words, whether the coefficient has changed over time, can be examined by testing whether α_3 is statistically different from 0. We can test for a similar change in the intercept across the two regimes by examining α_1 .

Table 3 shows the coefficient estimates for the regression model specified in equation 12.

Table 3. Coefficients Estimates Assuming a Discrete Change in the ELECTRONICVEHICLE Coefficient			
	Coefficient Estimate	Standard Error	p-Value
α_0 : INTERCEPT	-2162.72	290.05	0.00
α_1 : Z	-41.82	76.59	0.59
α_2 : ELECTRONICVEHICLE	-1350.47	295.38	0.00
α_3 : Z*ELECTRONICVEHICLE	501.10	281.10	0.08
β_2 : VEHICLECONDITION	727.24	68.63	0.00
β_3 : VEHICLEVALUATION	0.95	0.01	0.00
β_4 : NUMBERBUYERS	3.50	1.89	0.07
$R^2 = 0.98.$			

Note that the estimate for α_3 is positive and significant, indicating that the ELECTRONICVEHICLE coefficient differed between the Early and Late regimes. This is consistent with the other tests.

A limitation of this method is that it only permits analysis of whether the intercept and the ELECTRONICVEHICLE coefficient changed over time. The other coefficients are assumed to be invariant across the two regimes. This assumption may be overly restrictive. Consider that each regime might have its own regression model to describe it, as shown in equations 13 and 14.

$$y_{\text{early}} = \beta_{0,\text{early}} + \beta_{1,\text{early}} x_{1,\text{early}} + \dots + \beta_{k,\text{early}} x_{k,\text{early}} + \varepsilon \quad (\text{eq. 13})$$

$$y_{late} = \beta_{0,late} + \beta_{1,late}x_{1,late} + \dots + \beta_{k,late}x_{k,late} + \varepsilon \tag{eq. 14}$$

The coefficients for each of the independent variables might differ across regimes. To facilitate testing this, we can combine these two models as shown in equation 15.

$$y = d_{early} * (\beta_{0,early} + \beta_{1,early}x_{1,early} + \beta_{2,early}x_{2,early} + \dots + \beta_{k,early}x_{k,early} + \varepsilon) + d_{late} * (\beta_{0,late} + \beta_{1,late}x_{1,late} + \beta_{2,late}x_{2,late} + \dots + \beta_{k,late}x_{k,late} + \varepsilon) \tag{eq. 15}$$

The d_{early} and d_{late} terms in equation 15 represent dummy variables. d_{early} is set to 1 for observations in the Early regime, and 0 otherwise. d_{late} is set up analogously. If an observation belongs to the Early regime (i.e., $d_{early} = 1$ and $d_{late} = 0$), then equation 15 reduces to equation 13, and coefficient estimates are given by $\beta_{1,early}$, $\beta_{2,early}$, etc. The parallel is true for observations in the Late regime. The advantage of estimating the coefficients for both regimes simultaneously is that linear hypotheses can be used to determine if coefficients differ across regimes. For example, it is straightforward to test whether $\beta_{1,early} - \beta_{1,late} = 0$ using common statistical software. For example, after fitting a regression model using STATA, the command for testing linear relationships among coefficients is “test.”

Table 4 shows the coefficients that result from applying this type of model to our empirical context.

	Early Regime		Late Regime	
	Coefficient Estimate (Std Error)	p-Value	Coefficient Estimate (Std Error)	p-Value
β_0 : INTERCEPT	-9381.80 (451.37)	0.00	-1116.57 (160.72)	0.00
β_1 : ELECTRONICVEHICLE	-1270.96 (110.52)	0.00	-975.64 (98.52)	0.00
β_2 : CONDITIONNUMBER	739.13 (42.13)	0.00	764.09 (46.61)	0.00
β_3 : VALUATION	0.94 (0.01)	0.00	0.94 (0.01)	0.00
β_4 : NUMBERBUYERS	1.53 (1.39)	0.27	3.10 (0.76)	0.00
$R^2 = 0.99$.				
Linear hypothesis test to determine if ELECTRONICVEHICLE (Early) - ELECTRONICVEHICLE (Late) = 0:				
F (1, 11048) = 3.98, p-Value = 0.05.				

A linear hypothesis test indicates that the ELECTRONICVEHICLE coefficient in the Early regime is statistically different from the ELECTRONICVEHICLE coefficient in the Late regime, which is consistent with the other tests.

A test for a discrete change in a coefficient is appropriate when the researcher has data from two or more discrete points in time. For example, if a researcher has a data set consisting of a group of observations from June 2003, another group from March 2004, and a third group from January 2005, a discrete change test can be a useful mechanism to determine if coefficient estimates vary across these time periods. If the researcher has data that spans a block of time more or less continuously (as opposed to being bunched at specific points), then methods designed for continuous coefficient evolution can be used.

An Integrated Strategy for Using These Methods

A strategy for using the methods described above is to apply them in concert. This can help a researcher to leverage the strengths of each method to investigate whether and how coefficients vary over time. For example, an initial rolling regression procedure can provide a general picture of how each coefficient might vary. The researcher can then use the plot of a coefficient against time to help determine the functional form of a process function for that coefficient. As another example, the researcher might perform a rolling regression using a relatively large step size. The resulting plot of coefficients against time can provide clues as to whether there are any structural breaks within the data set, which can then be further examined via a test for a discrete change in the coefficient. This could

potentially illuminate external factors that trigger changes in the coefficients, such as new legislation or a system upgrade. Using multiple methods in concert allows a researcher to capitalize on each of their strengths and should increase confidence in any effects observed.

Implications and Conclusions

The data sets used by IS researchers are often structurally different than many data sets used in other fields, which has implications for investigating time-varying coefficients. Many of the most popular models for investigating time-based effects, such as AR (autoregressive), MA (moving average), and ARCH (autoregressive conditional heteroskedasticity) models, are designed for time series or panel data and are not appropriate for the pooled cross-sectional data sets often used in IS research. Our objective has been to highlight the peculiarity of the structure of many IS data sets and to present existing methods that are appropriate for investigating time-varying coefficients in these data sets. The arguments in this paper are by no means unique to the IS field, as other disciplines use these methods to investigate time-based effects in data sets that are not structured as classic time series or panel data. For example, see Beck (1983) for a survey of techniques appropriate for the political science field. However, there are two reasons why we believe this paper should be of special interest to IS researchers. First, relationships of interest in IS research are inherently dynamic due to changes in technology and how people and firms use technology. Thus, we should expect our coefficients to change over time, and we should be knowledgeable of methods to detect those changes. Second, IS data sets are often not structured as panels or classic time series, so many of the most well-known models are not appropriate. We have endeavored to highlight an alternative set of models, suitable for the types of pooled cross-sectional data that many IS researchers analyze.

We have used an e-commerce example in this paper, but these methods are not specific to e-commerce. They apply to any data set in which different units are measured at different points in time. For example, consider a researcher who conducts multiple surveys at different points in time to study IS adoption and usage. It may be unrealistic to survey the same respondents each time. Thus, this data will be structured as a pooled cross-section rather than as a panel, and the methods described herein will be useful if that researcher seeks to investigate how relationships affecting IS usage change over time.

As the IS field continues to evolve, we expect a common set of statistical techniques to emerge as tools for the IS researcher's toolkit. This has been the case in other, older, disciplines. For example, PhD students in economics learn not only the tools of econometrics but also the tools of individual decision theory and game theory. PhD students in operations research often learn linear programming techniques. Many marketing PhD programs have a PhD seminar devoted to "marketing models," which are a set of statistical models particularly well suited for marketing research topics. For example, many of these models pertain to customer choice among product alternatives and are variations of logit models. It is important to note that not all economists use game theory, not all operations researchers use linear programming, and not all marketing scholars analyze consumer choice models. However, they are all usually trained in these techniques, as the techniques are part of what defines their disciplines. Perhaps as the IS field continues to evolve, we will amass a series of "IS models" that will become standard material for IS research training. We suggest that an understanding of the methods available for modeling changes in relationships over time should belong in every IS researcher's toolkit. This paper represents a step in that direction. After all, the subject matter of IS is inherently dynamic, and we should embrace and exploit this dynamism in our work. This is one way we can further define our discipline.

To illustrate the methods discussed in this paper, we tested for time-varying coefficients in a data set gathered from the wholesale automotive market. We observed that a new technologically based product presentation mechanism originally caused buyers to pay less for vehicles, but that this effect lessened over time, perhaps as buyers grew more comfortable with this mechanism. By investigating whether coefficients varied over time, we were able to examine how new technologies are affecting market practice in one empirical context.

Changes in market practice are but one example of how information technology affects both business and society. Analysis of time-varying coefficients is one way we can observe the speed and trajectory of information technology's impact, which is a topic of central interest to our discipline.

References

- Andrews, D. W. K. "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* (61:4), July 1993, pp 821-856.
- Ba, S., and Pavlou, P. "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior," *MIS Quarterly* (26:3), September 2002, pp 243-268.
- Banker, R., and Mitra, S. "Impact of Information Technology on Agricultural Commodity Auctions in India," International Conference on Information Systems, Association for Information Systems, Las Vegas, NV, 2005, pp. 97-107.
- Beck, N. "Time-Varying Parameter Regression Models," *American Journal of Political Science* (27:3), August 1983, pp 557-600.
- Brown, R. L., Durbin, J., and Evans, J. M. "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society (Series B)* (37:2) 1975, pp 149-192.
- Chatfield, C. *The Analysis of Time Series: An Introduction*, Chapman & Hall/CRC, Boca Raton, FL, 1996.
- Chen, P. Y., and Hitt, L. M. "Measuring Switching Costs and the Determinants of Customer Retention in Internet-Enabled Businesses: A Study of the Online Brokerage Industry," *Information Systems Research* (13:3), September 2002, pp 255-274.
- Chow, G. C. "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica* (28:3), July 1960, pp 591-605.
- Dellarocas, C. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* (49:10), October 2003, pp 1407-1424.
- Devaraj, S., and Kohli, R. "Performance Impacts of Information Technology: Is Actual Usage the Missing Link?," *Management Science* (49:3), March 2003, pp 273-289.
- Enders, W. *Applied Econometric Time Series*, John Wiley & Sons, Inc., Hoboken, NJ, 2004.
- Farley, J., and Hinich, M. "Testing for a Shifting Slope Coefficient in a Linear Model," *Journal of the American Statistical Association* (65:331), September 1970, pp 1320-1329.
- Fichman, R. G., and Kemerer, C. F. "The Assimilation of Software Process Innovations: An Organizational Learning Perspective," *Management Science* (43:10), October 1997, pp 1345-1363.
- Gatignon, H., and Hanssens, D. M. "Modeling Marketing Interactions with Application to Salesforce Effectiveness," *Journal of Marketing Research* (24:3), August 1987, pp 247-257.
- Ghose, A., Smith, M. D., and Telang, R. "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research* (17:1), March 2006, pp 3-19.
- Hamilton, J. D. *Time Series Analysis*, Princeton University Press, Princeton, NJ, 1994.
- Hinich, M., and Roll, R. "Measuring Nonstationarity in the Parameters of the Market Model," *Research in Finance* (3), January 1981, pp 1-51.
- Jiang, Z., and Benbasat, I. "Virtual Product Experience: Effects of Visual and Functional Control of Products on Perceived Diagnosticity and Flow in Electronic Shopping," *Journal of Management Information Systems* (21:3), Winter 2004, pp 111-147.
- Kauffman, R. J., and Wood, C. A. "Analyzing Competition and Collusion Strategies in Electronic Marketplaces with Information Asymmetry," working paper, 2000.
- Kauffman, R. J., and Wood, C. A. "The Effects of Shilling on Final Bid Prices in Online Auctions," *Electronic Commerce Research and Applications* (4:1), Spring 2005, pp 21-34.
- Lee, H. G. "Do Electronic Marketplaces Lower the Price of Goods?," *Communications of the ACM* (41:1), January 1998, pp 73-80.
- Naik, P. A., and Raman, K. "Understanding the Impact of Synergy in Multimedia Communications," *Journal of Marketing Research* (40:4), November 2003, pp 375-388.
- Overby, E., and Jap, S. "Electronic vs. Physical Market Mechanisms: Testing Multiple Theories in the Wholesale Automotive Market," working paper, 2006.
- Wildt, A. R., and Winer, R. S. "Modeling and Estimation in Changing Market Environments," *Journal of Business* (56:3), July 1983, pp 365-388.
- Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA, 2002.