Continuous Data Level Auditing: Business Process Based Analytic Procedures in an Unconstrained Data Environment[1]

Michael G. Alles
Alexander Kogan
Miklos A. Vasarhelyi

Department of Accounting & Information Systems
Rutgers Business School
Rutgers University
180 University Ave
Newark, NJ 07102

Jia Wu

Department of Accounting and Finance
University of Massachusetts – Dartmouth
285 Oldwestport Road
North Dartmouth, MA 02747

November 22, 2006

[1] Comments are welcome and may be addressed to Alexander Kogan at kogan@rbsmail.rutgers.edu. We thank the KPMG CA/R/Lab at Rutgers Business School for support. We also thank participants, and especially the discussant, Tony Tinker, at the Four Schools Conference at Baruch College for helpful feedback, as well as Carol Brown and participants at the 2005 and 2006 American Accounting Association Meetings. All views expressed in this paper are those of the authors alone.

Continuous Data Level Auditing: Business Process Based Analytic Procedures in an Unconstrained Data Environment

Abstract:

This paper designs a Continuous Data Level Auditing system utilizing business process based analytical procedures in a setting of unconstrained data availability and evaluates the system's performance using real-world data sets extracted from the supply chain data warehouse of a large healthcare management firm. The first component of the proposed CA system utilizes automatic transaction verification to filter out exceptions, which are transactions violating formal business process rules. The second component of the system creates business process audit benchmarks which we denote as "Continuity Equations", and define as stable *probabilistic models of highly disaggregated business processes* — the expectation models for process based analytical procedures. Our first objective is to investigate three probabilistic models that can serve as candidates for our continuity equation benchmarks: a Simultaneous Equation Model, a Vector Autoregressive model and a Linear Regression Model. Our second objective is to take advantage of the fundamental characteristic of a continuous audit system that assurance takes place close to the transaction date to design a set of online learning and error correction protocols for automatic model inference and updating. Our third objective is to examine the impact of the choice of the level of data aggregation that unconstrained data availability makes possible on anomaly detection performance. Using a seeded error simulation approach, we find that under most circumstances the use of real time error correction results in superior performance. We also find that each candidate audit benchmark model has its own strengths and weaknesses, and hence recommend that different continuity equation models can complement one another and be used concurrently in analytic procedures. Overall, our results indicate that when auditors have access to unconstrained data, the richness of that disaggregated data combined with the ability to make real time error correction makes error detection robust across a variety of expectations models, a key conclusion in support continuous data level auditing.

Keywords: continuous auditing, analytical procedures, error correction.
Data availability: The data is proprietary. Please contact the authors for details.

I. Introduction

Continuous Auditing with Unconstrained Data Availability

Business is in the process of a fundamental transformation towards the digital economy (Vasarhelyi and Greenstein, 2003). With companies having implemented networked integrated Enterprise Resource Planning (ERP) systems (such as SAP R/3, Oracle Applications, PeopleSoft) as their basic information infrastructure, management and control of organizations is shifting to a data-centric, process-oriented paradigm.[2] The requirements of Section 404 of the Sarbanes-Oxley Act for rigorous controls over financial reporting also focus attention on how data is processed and used within the company, while the mining of customer and operational data is essential for company's pursuing strategies of total customer satisfaction and total quality control.

In response to these fundamental changes in the business environment, public accounting firms and internal auditing departments are now facing the opportunities and challenges associated with the development and deployment of *continuous auditing* (CA) which is largely automated data intensive audit procedures with decreased latency between the transaction event and the provision of assurance.[3] In the limit, the auditor would access real time streams of the entire universe of the company's transactions rather than being restricted to a small sample gathered at a single moment of time (as in the annual stock take, to take the prototypical example). The feasibility of creating such a real time, automated audit methodology arises from the capability of the company's ERP system to make available to auditors business data of far finer granularity in time and detail than has ever been cost effectively accessible before.[4]

Continuous auditing is becoming an increasingly important area in accounting, both in practice and in research, with conferences held around the world heavily attended by both academics and practitioners.[5] The major public accounting firms all have CA initiatives under way, and major software vendors are also now aggressively developing and marketing

---

[2] Vasarhelyi and Greenstein, 2003.

[3] The CICA/AICPA report defines CA as *"a methodology that enables independent auditors to provide written assurance on a subject matter using a series of auditors' reports issued simultaneously with, or a short period of time after, the occurrence of events underlying the subject matter."* http://www.cica.ca/index.cfm/ci_id/989/la_id/1.htm

[4] Vasarhelyi et al 2004.

[5] Alles et al 2006b.

CA software solutions (with SAP's acquisition of Virsa being a recent example). PricewaterhouseCoopers (2006) in their recent survey state that *"Eighty-one percent of 392 companies responding to questions about continuous auditing reported that they either had a continuous auditing or monitoring process in place or were planning to develop one. From 2005 to 2006, the percentage of survey respondents saying they have some form of continuous auditing or monitoring process within their internal audit functions increased from 35% to 50%—a significant gain."*[6] On the research front, the ongoing survey of the CA research literature by Brown et al. (2006) lists at least 60 papers in the area, ranging from behavioral research, to system design and analytical models and implementation case studies.

Notably, however, there is a dearth of empirical research or case studies of new CA methodological developments due to the lack of data availability and difficulties of access to companies implementing CA. As a consequence, what is missing from both the academic and professional literature is a rigorous examination of how CA will impact the day to day practice of auditing, and in particular, how auditing will cope with the shift from data scarcity to data wealth, from periodic, archival to real time streaming data. This is a critical omission since much of existing audit practice is driven precisely by lack of data and the cost of accessing it: hence auditors do sampling, establish materiality thresholds for investigations and carry out analytic procedures before substantive testing so that they can focus only on likely trouble spots. Will any of these familiar practices survive in an age of digital firms with trivial costs of data storage, access and communication? Or to put it another way, can auditors afford to keep doing what are effectively short cuts, the justification for which has long since become obsolete?

In our opinion, it is one thing for an auditor to choose to base audit procedures on limited data when data is very costly to obtain, and quite another to continue doing so when unconstrained data is readily available. It is the latter situation that the audit profession will increasingly find itself in until auditing procedures and systems are developed that can exploit the availability of timely and highly disaggregated data. In other words, the audit profession has to either answer the question of what it plans to do with all the data that it will soon be able to easily obtain, data which provides the level of detail an order of

[6] CFO.com, June 26, 2006.

magnitude beyond the sampled, highly aggregated data that is the basis of much of the current audit methodology—or else, to explain why data is being thrown away unused.

Certainly there are challenges in dealing with this quantity of data whose quality is variable, but the costs facing the auditor in processing these data have to be weighted against the potential benefits associated with the opportunity to free ride off the multi-billion dollar investment made by the firms in implementing the ERP systems, which make the provision of that data possible in the first place. It is incumbent on the auditors to develop methodologies that exploit that opportunity so that they can provide their clients with higher quality, more effective and efficient audits.

Business Process Based Analytic Procedures

Auditing is defined as *"a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users."*[7] Thus the scope of auditing is driven not only by what evidence is available, but also whether there exist benchmarks—the "established criteria"—to compare that audit evidence against. Those benchmarks provide guidance about what the data is supposed to look like when drawn from a firm operating without any anomalies.

One of the key roles played by benchmarks in modern auditing is in the implementation of Analytic Procedures (AP), which Statement on Auditing Standards (SAS) No. 56 defines as the "*evaluation of financial information made by a study of plausible relationships among both financial and nonfinancial data*". SAS 56 requires that analytic procedures should be performed during the planning and review stages of an audit, and recommends their use in substantive testing in order to minimize the subsequent testing of details to areas of detected concern. That sequence is dictated because manually undertaken tests of detail are so costly that they are resorted to only if the account balanced based AP tests indicate that there might be a problem. Both the timing and nature of standard analytic procedures are thus brought into question in a largely automated continuous auditing system with unconstrained data.

[7] Auditing Concepts Committee (1972, page 18).

The hypothesis driving our research is that making use of that unconstrained data it is possible to design analytic procedures which have an unprecedented degree of correspondence to underlying business processes. Business processes (BP), are defined (Davenport and Short 1990) as *"a set of logically related tasks performed to achieve a defined business outcome"*. As the large investments in ERP systems indicate, business processes are considered today to be the fundamental atomic elements that make up a company, as much as the listing of its physical assets or employees might have been in earlier eras.[8]

Creating business process based benchmarks requires unconstrained data at a highly disaggregate level, far below the level of account balances that are used in most analytic procedures today, such as ratio analysis or comparisons with prior year financials. Testing the content of a firm's data flow against such benchmarks focuses on examining both exceptional transactions and exceptional outcomes of expected transactions. Ideally CA software will continuously and automatically monitor company transactions, comparing their generic characteristics to observed/expected benchmarks, thus identifying anomalous situations. When significant discrepancies occur, alarms will be triggered and routed to the appropriate stakeholders.

The objective of this project then is to explore the benefits of using business process based analytic procedures to create a system of continuous data level auditing.

Continuity Equations

The first component of the proposed CA system utilizes automatic transaction verification to filter out exceptions, which are transactions violating formal BP rules. The second component of the system creates business process audit benchmarks which we denote as **Continuity Equations** (CE), and define as stable *probabilistic models of highly disaggregated business processes,* as the expectation models for process based analytical procedures. Continuity Equations are commonly used in physics as mathematical expressions of various conservation laws, such as the law of the conservation of mass.[9] In the continuity equation metaphor, each business process is analogous to a control volume made up of a variety of transaction flows, or business activities. If transaction flows into and

[8] Porter 1996.

[9] For a control volume that has a single inlet and a single outlet, the principle of conservation of mass states that, for steady-state flow, the mass flow rate into the volume must equal the mass flow rate out.

out of each BP are equal, the business process would be in a steady-state, free from anomalies. Otherwise, if spikes occur in the transaction flows, the steady-state of the business process cannot be maintained.[10]

The object of having audit benchmarks consisting of CEs is to capture the dynamics of the fundamental business processes of a firm, but since those processes are unobservable, in practice the CE is a data driven statistical estimate. Which probabilistic representation of those underlying business processes provides the CE that is most effective as an audit benchmark is an empirical issue. Once identified, CEs are applied to the transaction stream to detect statistical anomalies possibly indicating business process problems. BP metrics used in CEs can be both traditional financial metrics (such as the dollar amounts of daily purchases) which are commonly used in auditing, and physical metrics (such as the quantity of items ordered, or the number of purchase orders placed) which are more common in engineering and statistical process quality control.

For CE based APs to be adequate to meet the needs of the CA environment, they have to satisfy a number of important criteria:

1. First, the process of creating continuity equations to serve as audit benchmarks in the AP tests should be largely automated and the CE models should be self-adaptive, requiring as little human intervention as possible. It is simply not feasible for auditors to manually select expectations models when dealing with highly disaggregated data without succumbing to data overload. Further, new data are continuously fed into the CA system through the firm's ERP systems. An AP model for CA must be able to assimilate additional information contained in the new data feeds, and adapt its CE benchmarks continuously.
2. Second, given that the ultimate objective for auditors applying AP is to detect anomalies and then to perform detailed testing to resolve these detected anomalies, AP models should be able to effectively and efficiently detect errors. Since highly disaggregated data is rarely examined in traditional audit practice, and certainly was not examined at the depth that we do in this

[10] The familiar accounting equation linking the balance sheet to the income statement—assets = liabilities + owners equity—is an example of a continuity equation, and verifying that it holds is one of the central tasks of the financial audit.

project, the measure of success for new CA-based AP tests is open to debate. There is no obvious "horse race" to be run, comparing the new versus the traditional AP procedures. Besides, we believe that traditional AP would still be used in the planning and final review stages of an audit even when CA is widely implemented. Hence, we focus on internal validity, examining whether AP tests that use highly disaggregated data are effective in detecting inherent and seeded errors in that data. Anomaly detection is why auditors utilize AP tests in the first place, and demonstrating effectiveness at this task sufficiently justifies our research initiative.

3. Further, to improve error detection capability, the AP models should be utilized to correct any detected errors as soon as possible to ensure that the subsequent predictions are based on the correct data as opposed to the erroneous ones.

Having developed this theoretical framework, we take the important next step of validating the proposed CA system design using a large data set of the supply chain procurement cycle data provided by a large healthcare management firm. This allows us to examine what form the CEs will take in a real world setting and how effective they are in detecting errors.

While the data is not analyzed in real time, the extent of the data we use mimics what a working data level CA system would have to deal with and it provides a unique testing ground to examine how audit procedures will adapt to deal with unconstrained data availability. Indeed, even when CA is ubiquitous, it is unlikely that the audit system will be able to "tap" the ERP data warehouse in real time. In order to maintain the integrity of the data base and to avoid slowing down priority operational access to it, the data will be fed in batches to the CA system. But even daily downloads undertaken overnight will still provide a far reduced latency between transaction and assurance than anything available today. To that extent, the fact that our project uses archived as opposed to streaming data is irrelevant because the key is that we have access to any raw data that we wish on any process that want to model at whatever level of detail that is deemed appropriate in order to create CE based audit benchmarks and undertake the AP tests that will characterize continuous data level auditing.

The unconstrained data is used to identify three key business processes in the procurement cycle: the ordering process, the receiving process, and the voucher payment process. The CE models of these three processes are estimated using the statistical methodologies of linear regression, simultaneous equation modeling, and vector autoregressive models. We design a set of online learning and error correction protocols for automatic model inference and updating. We use a seeded error simulation study to compare the anomaly detection capability of the discussed models. We find that under most circumstances the use of real time error correction results in superior performance. We also find that each type of CE models has its strengths and weaknesses in terms of anomaly detection. These models can be used concurrently in a CA system to complement one another. Finally, we demonstrate that the use of disaggregated data in CE can lead to better anomaly detection when the seeded errors are concentrated, while yielding no improvement when the seeded errors are dispersed.

The remainder of this paper is organized as follows. Section 2 provides a review of the relevant literatures in auditing, CA, business processes and AP upon which our work is based and to which the results of the paper contribute. Section 3 describes the design and implementation of data-oriented CA systems, Section 4 discusses the critical decision choice of how to aggregate the transactional data, and the CE model construction using three different statistical methods, with Section 5 comparing the ability of the CE-based AP tests in detecting anomalies under various settings. Section 6 discusses the results, identifies the limitations of the study, and suggests future research directions in this domain. Section 7 offers concluding comments.

II. Literature Review

This paper draws from and contributes to multiple streams of literature in system design, continuous auditing and analytical procedures.

Continuous Auditing

The seminal papers on continuous auditing are Groomer and Murthy (1989) and Vasarhelyi and Halper (1991). They pioneered the two modern approaches towards designing the architecture of a CA system, the embedded audit modules and the control and monitoring layer, respectively. The literature on CA since then has increased considerably, ranging from the technical aspects of CA (Kogan et al. 1999; Woodroof and Searcy 2001;

Rezaee et al. 2002; Murthy 2004; Murthy and Groomer 2004, etc.) to the examinations of the economic drivers of CA and their potential impact on audit practice (Alles et al. 2002 and 2004; Elliott 2002; Vasarhelyi 2002; Searcy et al. 2004). Kogan et al. (1999) propose a program of research in CA. In the discussion of the CA system architecture they identify a tradeoff in CA between auditing the enterprise system versus auditing enterprise data. A recent study by Alles et al. (2006) develops the architecture of a CA system for the environment of highly automated and integrated enterprise system processes, and shows that a CA system for such environments can be successfully implemented on the basis of continuous monitoring of business process control settings. This paper focuses on the enterprise environment in which many business processes are not automated and their integration is lacking, and proposes to design a CA system architecture based on data-oriented procedures. In this development, it utilizes the approach of Vasarhelyi et al. (2004) that introduces four levels of CA assurance having different objectives. More specifically, this paper develops a CA methodology for the first and third CA levels: transaction verification and assurance of higher-level measurements and aggregates.[11]

The unavailability of data to researchers is the likely cause of a lack of empirical and case studies on CA in general and on analytical procedures for CA in particular. This paper contributes to the CA literature by providing empirical evidence to illustrate the advantages of CA in real-time problem resolution. More specifically, we show that potential problems can be detected in a more timely fashion, at the transaction stream level as opposed to the account balance level. Traditionally, analytical procedures are applied at the account balance level after the business transactions have been aggregated into account balances. This would not only delay the detection of potential problems but also create an additional layer of difficulty for problem resolution due to a large number of transactions that are aggregated into accounting numbers. The focus on auditing the underlying business processes alleviates this problem by utilizing much more disaggregated information in continuous auditing.

Enterprise Business Process Modeling and Continuity Equations

In his seminal papers, McCarthy (1979 and 1982) proposes the Resource-Event-Agent (REA) paradigm for representing accounting objects in a shared data environment. Geerts and McCarthy (1997, 2005) extend and refine the REA accounting model to engineer

[11] The other two levels are compliance and judgment verification.

business processes and tasks by decomposing the enterprise value chain into a number of interrelated of business processes. Those processes that involve inflows and outflows of economic resources are called economic processes. Each economic process contains at least two related economic events: "a decrement event that consumes the inputted resource and an increment event that acquires the outputted resource."

We follow the REA approach which constructs the enterprise business model starting with the economic processes. We develop a CA data-oriented methodology around the key economic processes and the economic resources, and illustrate this methodology on the example of the supply chain of the firm. This approach can also be viewed as an extension to CA of the modern business process auditing approach proposed by Bell et al. (1997). They advocate a holistic approach to auditing an enterprise: structurally dividing a business organization into various business processes (e.g. the revenue cycle, procurement cycle, payroll cycle, and etc.) for the auditing purpose. They suggest the expansion of the focus of auditing from business transactions to the routine activities associated with different business processes.

Building on the REA perspective, Vasarhelyi and Halper (1991) are the pioneers to taking advantage of online technology and modern networking to develop a procedure for continuous auditing. Their study introduces the concept of continuous analytical monitoring of business processes, and discusses the use of key operational metrics and analytics to help internal auditors monitor and control AT&T's billing system. They use the operational process auditing approach and emphasize the use of metrics and analytics in continuous auditing. This is the first study to adopt the term "Continuity Equations" which is used in modelling of how billing data flowed through business processes and accounting systems at AT&T. The choice of the expression by Vasarhelyi and Halper is clearly driven by the fact that as with the conservation laws in physics, in a properly functioning accounting control system there should not be any "leakages" from the transaction flow.

This paper continues the application of the concept of CE to model the relationships between the metrics of key business processes that make up the REA view of the firm, while building on the original implementation of CA as described in Vasarhelyi and Halper (1991). The broader implications of their model have been obscured in the subsequent focus of the CA literature on technology enablers and the frequency of reporting. But in a very real sense, attaining the full potential of CA requires the utilization of not only of its well known capability of decreasing audit latency, but also of taking advantage of unconstrained data availability to create audit benchmarks that are not only timelier but provide a more accurate, detailed and dynamic benchmark of fundamental business processes. In that sense CE based continuous auditing brings the worlds of auditing and REA into closer alignment.

Analytical Procedures

Analytic procedures reduce the audit workload and cut the audit cost because they help auditors focus substantive tests of detail on material discrepancies. In applying analytical procedures an auditor first develops an expectations model to make a prediction about the value of an important business metric such as an account balance. Then, the auditor compares the predicted value with the actual value of the metric. Finally, if the variance between the two values exceeds a pre-established threshold, an alarm is triggered leading to further investigation by the auditor of the discovered anomaly.

There are extensive research studies on analytical procedures in auditing. Many papers discuss various analytical procedures ranging from financial ratio analysis to linear regression modeling that focus on highly aggregated data such as account balances (Hylas and Ashton 1982; Kinney 1987; Loebbecke and Steinbart 1987; Biggs et al. 1988; Wright and Ashton 1989; Hirst and Koonce, 1996). The percentages of errors found using such analytical procedures are usually not high, varying between 15% and 50%. Only a few papers examine analytical procedures for more disaggregated data. Dzeng (1994) compares 8 univariate and multivariate AP models using quarterly and monthly financial and non-financial data of a university, and concludes that disaggregated data yields better precisions in a multivariate time-series based expectations model. Other studies also find that applying AP models to higher frequency monthly data can improve analytical procedure effectiveness (Chen and Leitch 1998 and 1999, Leitch and Chen 2003). By contrast, Allen et al. (1999) use both financial and non-financial monthly data of a multi-location firm and do not find any supporting evidence that geographically disaggregate data can improve analytical procedures.

In this study we build a data level CA system which utilizes CE based analytical procedures applied to even more highly disaggregate daily metrics of business processes. We investigate several different probabilistic models of those business processes to serve as our CE based audit benchmark: the Simultaneous Equation Model (SEM), the Vector

Autoregressive Models (VAR) and the Linear Regression Model (LRM). The use of SEM in analytical procedures has been examined by Leitch and Chen (2003), but only using monthly financial statement data. Their finding indicates that SEM can generally outperform other AP models including Martingale and ARIMA.

As far we can ascertain, the Vector Autoregressive Model has not been fully explored in the auditing literature. There are a number of studies utilizing univariate time series models (Knechel 1988; Lorek et al. 1992; Chen and Leitch 1998; Leitch and Chen 2003), but only one, by Dzeng (1994), which uses VAR. Dzeng concludes that VAR is better than other modeling techniques in generating expectation models, and he specifically recommends using Bayesian VAR (BVAR) models. The computational complexity of VAR used to hamper its application as an AP model, since appropriate statistical tools were not readily available in the past. However, the recent developments in statistical software facilitate the application of this sophisticated model.[12] The VAR model can not only represent the interrelationships between BPs but also capture their time series properties. Although (to the best of our knowledge) VAR has been discussed only once in the auditing literature, studies in other disciplines have either employed or discussed VAR as a forecasting method (Swanson 1998; Pandher 2002).

III. Design and Implementation of a Continuous Data Level Auditing System

The objective of a CA system designed in this study is to provide close to real-time assurance on the integrity of certain enterprise business processes. As in conventional auditing, such a system can adopt either one of two different types of procedures: those monitoring business process controls and those analyzing business process transactions. As Alles et al. (2006) indicate, business process control monitoring requires that the client possesses a modern integrated IT infrastructure, and faces challenges even then. They also show that even today few firms have the kind of tight, end to end data integration that continuous control monitoring depends upon. This paper focuses on designing a CA system for the much more common enterprise environments in which data is derived from multiple legacy systems that lack centralized and automated controls. This lack of a control based

[12] Starting with version 8, SAS (Statistical Analysis System) allows users to make multivariate time series forecasts using a very sophisticated VARMAX procedure.

monitoring system is why the proposed CA system is data-oriented instead, and the provision of assurance based on verifying transactions and business process based analytic procedures.

The architecture of the designed data level CA system is driven by the procedures it has to implement. While the subject matter it deals with is quite different from that in conventional auditing, one can view its procedures as analogous to automated substantive audit procedures, including transaction testing and analytical procedures. Therefore, the two main components of the CA system are those executing automatic transaction verification and testing using CE based automatic analytical processes.

[Insert Figure 2 here]

The implementation of the transaction verification component of the CA system is based on identifying business process rules and formalizing them as transaction integrity and validity constraints. Every recorded transaction is then checked against all the formal rules in the component, and if it violates any of the rules, then the transaction is flagged as an *exception*. Every exception generates a CA alarm in the CA system, which it sends to the appropriate parties for resolution. Since the alarm specifies which formal business process rules are violated by the exception, resolving exceptions should be a fairly straightforward task. Once the transaction data is verified it is in an acceptable form to be used to develop the CE based audit benchmarks for AP.

Unconstrained Data Provision

Enterprise systems that support key business processes routinely collect business process data in the unfiltered highly disaggregated form. If the enterprise has implemented an integrated ERP system, then BP data is readily available in the ERP central database. However, the most common current situation is that the enterprise system landscape consists of a patchwork of different systems, many of which are legacy ones and are often file-based. In such enterprise systems direct real-time access to business process data is highly problematic, if at all possible at any reasonable expense of time and effort. Therefore, a data-oriented CA system usually cannot be cost-effectively deployed in such environment unless the enterprise deploys an overlay data repository commonly known as "Business Data Warehouse". This is a relational database management system specially designed to host business process data provided by the other enterprise systems, including the legacy cycle-focused ones (such as sales processing or accounts receivable). While the main functionality of a data warehouse is online analytical processing, the CA system developed here relies only on its function as the global repository of business process data. The availability of unconstrained business process data, meaning that the auditor can access any raw, unfiltered and disaggregated data that is required for the construction and operation of CE based AP tests is the critical enabler of the proposed CA system.

Data Description

Our simulated implementation of the data-oriented CA system focuses on the procurement-related business processes and utilizes the data sets extracted from the data warehouse of a healthcare management firm with multi-billions of dollars in assets and over two million employees. The firm is a major national provider of healthcare services, with a network composed of locally managed facilities that include numerous hospitals and outpatient surgery centers all over the US and overseas. A key strategic driver for the firm is the management of their supply chain which provides everything from paper towels to heart/lung machines to their various operating units through dozens of warehouses spread throughout the United States.

We were approached by the firm's Internal Audit in 2003 to consider how to improve the assurance they could provide over their supply chain, focusing on sample of warehouses in one region of the US. What they were willing to provide was extracts from their transactional database, which while only a sample limited in time and geography, still encompassed megabytes of data, several orders of magnitude more detailed than anything typically examined in a standard audit.

The data sets include all procurement cycle daily transactions from October 1st, 2003 through June 30th, 2004. The number of transaction records for each activity ranges from approximately 330,000 to 550,000. These transactions are performed by ten facilities of the firm including one regional warehouse and nine hospitals and surgical centers. The data was first collected by the ten facilities and then transferred to the central data warehouse in the firm's headquarters.

Transaction Verification

Following the BP auditing approach, as the first step, we identify the following three key business processes in the supply chain procurement cycle: ordering, receiving, and voucher payment, which involve six tables in our data sets. The economic resources involved include facility items (inventory) and voucher payments (cash).

[Insert Figure 2 here]

This data is uploaded to the data warehouse for the underlying legacy systems, which are lacking many automated controls present in modern ERP systems. Not surprisingly, then, there are numerous data integrity issues, which have to be identified by the transaction verification component of the CA system before the data is suitable for AP testing. To simulate the functionality of the transaction verification component, we formally specify various data validity, consistency, and referential integrity constraints, and then filter through them all the available transactions.

Two categories of erroneous records are removed from our data sets: those that violate data integrity and those that violate referential integrity. Data integrity violations include but are not limited to invalid purchase quantities, receiving quantities, and check numbers.[13] Referential integrity violations are largely caused by many unmatched records among different business processes. For example, a receiving transaction cannot be matched with any related ordering transaction. A payment for a purchase order cannot be matched with the related receiving transaction. Before we can build any analytical model, these erroneous records must be eliminated. This removal simulates the action of the transaction verification component of the CA system. Note that in a very tightly integrated enterprise environment such transactional problem would have been prevented by the client's ERP system.

An additional step in the transaction filtering phase is to delete non-business-day records. Though we find that sporadic transactions have occurred on some weekends and holidays, the number of these transactions accounts for only a small fraction of that on a working day. However, if we leave these non-business-day records in our sample, these records would inevitably trigger false alarms simply because of low transaction volume.

[13] We found negative or zero numbers in these values which can not always be justified by our data provider.

While the simulated verification of transactions relied on fairly straightforward business rules described above, the client firm considered that just the exceptions identified at this stage were a major source of value added from the project. It is to be anticipated that as legacy systems are gradually superseded by the firm's ERP system with stronger automated controls, the transaction verification component of the CA system will be catching fewer and fewer problems. Conversely, the fact that any were caught at all indicate the value of this limited form of automated continuous auditing, since these transaction level errors had escaped detection from the standard practices being employed by either the firm's internal or external auditors, practices which obviously were not designed to cope with data universality.

Business Process based Analytic Procedures

The transaction verification stage of continuous data level auditing is based on user specified rules designed to catch obvious errors in individual business events. Whatever value added this step may provide, catching such errors is hardly the primary purpose of auditing. Rather, the object of analytic procedures is the "*study of plausible relationships among both financial and nonfinancial data*" in order to detect anomalous patterns in the firm's performance and the way in which it is reported.

The implementation of the AP component of the CA system first requires creation of the process level continuity equations which can serve as benchmarks for the AP tests. Instead of testing transaction by transaction for obvious errors, the AP test contrasts the stream of data against a probabilistic model of how that data should look like if there were no untoward events happening in the firm. Such models usually take the form of statistically stable relationships between business process metrics (and possibly some exogenous factors). Every business process metric is calculated over a subset of transactions corresponding to intervals along some important business process dimensions (such as time, region, product, customer, etc). Since the relationship between the metrics holds only probabilistically, the model also has to specify the acceptable range of variation of the residuals. An anomaly arises when the observed values of the metrics results in residuals which fall outside this acceptable range. Every anomaly generates a CA alarm in the CA system, which it sends to the appropriate parties for resolution. In contrast with an exception, which is associated to an individual transaction, an anomaly is associated with a subset of transactions used to calculate the values of the metrics. Therefore, the resolution of an anomaly is not straightforward. Moreover, an anomaly is not necessarily indicative of a problem, since it can simply be due to statistical fluctuation.

Note that in the workflow of the proposed data-oriented CA system the first stage of processing is the verification of transactions. It has to be contrasted with the fact that in conventional auditing analytical procedures are used first to identify areas of concern, and then transaction testing is focused on the identified risky areas. The rationale for this sequence is that it makes it possible to reallocate the sample counts so as to increase either the effectiveness or the efficiency of substantive testing. Since in a CA system the verification of transactions are automatically performed on the entire population of data without the need for sampling. Therefore, the transaction verification component of the CA system processes every transaction in the business process stream, and serves as a filter screening out the identified exceptions. Then this filtered stream of transactions is further processed by the analytical procedures component of the system to ascertain the absence of anomalies. Thus, the CA system reverses the sequence of procedures of traditional auditing to capitalize on the capabilities of modern IT to verify the entire stream of business transactions against the formalized set of process rules.

The rationale for analytic procedures in the proposed CA architecture is two-fold. First, one can never assume that the set of formalized business process rules completely defines the set of constraints business transactions have to satisfy. Given this possibility, analytical procedures serve as the second line of defense. Anomalies identified by the CA system can signal the presence of likely abnormal transactions that are not discovered by the user defined rules of the transaction verification filter. This is an indication either of some exceptional event, such as fraud, or if the anomaly occurs often enough, of the need to update the rule set of the transaction verification system.

Second, analytical procedures can identify certain business process irregularities that cannot be caught in principle by the transaction verification component because they are not due to the violation of any business rules. At this stage of this research project into continuous data level assurance we are only investigating the efficacy of CE based AP tests in detecting generic errors, as we explain below. But once that is established, the next step is to develop the equivalent of the rules in the transaction verification stage, by identifying particular patterns at the data stage that correspond to particular areas of concern to the auditor. Obvious issues to examine include round tripping, channel stuffing and other well known cases of accounting fraud.

In this specific area of procurement, there is a possibility that a batch of purchase orders may be destroyed without any notification by the delivery mechanism external to the enterprise. In this case, the unexpected drop-off in item deliveries can be identified as anomalous by the analytical procedures, which will thus provide the precious early warning of a process problem. Another procurement-related example has to do with regular payments of vendor invoices to take advantage of early payment discounts. Then, the process rules may not require payments before the invoice due date, while the practice is to pay early. If, say, due to a staffing change in the accounts payable department, the early payments are not processed on time, there will be no violation of process rules, but the analytic procedures would still be able to signal an anomaly after identifying an unexpected decline in the number of payments processed. Human investigations of anomalies identified by the analytic procedures should be able to discern the root cause of the anomalies (if any) and initiate the necessary corrective actions.

Much more research is needed to create a library of anomalous business patterns that can be loaded into the CA system and serve to direct human auditors to particular problem areas. Before that can happen, we must first show that it is possible to create process level benchmarks, the continuity equations, in the first place.

IV Models of Continuity Equations

Following the BP auditing approach, we have identified three key business processes for our sample firm which include ordering, receiving, and voucher payment processes. The analytical procedures component of the CA system is based on benchmarks which model the interrelationships between these processes. A critical issue in modeling business processes analytically is the choice of BP metrics. The traditional accounting choice has been the use of financial measures (e.g., dollar amounts), driven in the first place by the reliance on ledger entries as the primary source of data. In an unconstrained data environment, however, modeling business processes can also be undertaken using use of other types of nonfinancial metrics such as physical measurements or document counts. The dollar amounts of each transaction or the number of transactions processed can also be used. In our study the transaction item quantity is selected as our BP metric. There is, of course, no conceptual reason why analytical procedures cannot utilize multiple metrics to examine transaction flows. Auditing on different metrics would enable auditors to detect a more diverse set of patterns of firm behavior.[14]

Once the BP metrics are chosen, the next step is to determine the appropriate degree of aggregation at which it is appropriate to conduct the AP test, and hence, the characteristics of the data used to construct the CE based benchmark.

Data Aggregation

The main argument against using aggregated data is that it inevitably leads to a loss of information about individual transactions. But aggregation can also make it possible to see the so called "big picture". The debate over how and to what extent to aggregate transactional data is as old as accounting itself, and its use of ledger accounts as a means of summarizing data. The key difference is that in an unconstrained data environment and with the technical ability to process such large data sets, the degree and nature of aggregation is now a choice that is open to auditors to make, rather than one forced on them by measurement constraints.

The main statistical argument for aggregation is that it can reduce the variability observed among individual transactions. For example, the transaction quantity can differ greatly among individual transactions, as well as the lag time between order and delivery, and delivery and payment. By aggregating the individual transactions, this variance can be significantly reduced, thus allowing more material anomalies to be detected more effectively. The fluctuations among individual transactions can also be smoothed by aggregating them, which facilitates the construction of a stable model. Otherwise, it would be infeasible to derive a stable model based on data sets with large variances because the model would either trigger too many alarms or lack the detection power. On the other hand, if individual

[14] We need to perform audits on different metrics besides financial numbers. For example, the Patriot Act requires that banks should report the source of money for any deposit larger than US$100,000 by its client. However, the mandatory reporting controls can be by passed by dividing the deposit over $100,000 into several smaller deposits. Even though the deposit amount each time is under the limit, the number of total deposits is over the limit. Auditors can only catch such fraudulent activity by using the number of deposit transactions as one of the audit metrics.

transactions are aggregated over a longer time period such as a week or a month, then the model would fail to detect many abnormal transactions because the abnormality would be mostly smoothed out by the longer time interval. Thus, the inescapable tradeoff that the more aggregated the metrics are, the more stable the analytical relationships are likely to be at a price of more missed detection. In the mean time, any anomaly involving a metric with higher level of aggregation, requires a more extensive (and expensive) investigation of the larger subpopulation of transactions if an alarm is triggered. Daily and weekly aggregations used in our analysis are natural units of time that should result in a reasonable trade-off between these two forces. Aggregation can be performed on other dimensions besides the time interval, and the choice of the aggregation levels has obviously to be made on a case by case basis considering the inherent characteristics of the underlying transactional data.

We use intermediate aggregates of transactions, such as aggregates of transactions of different units in the enterprise, aggregates of transactions with certain groups of customers or vendors. This is a novelty since traditional substantive testing is done either at the most disaggregated level, or at the most aggregated level. Substantive tests of transactions are done at the most disaggregated level of individual transactional data, but this is done in order to verify the correctness of that individual transaction rather than to gain a perspective of the overall business process. Tests of details of account balances are obviously applied at the most aggregated level. All standard analytical procedures are used for analyzing the account balances or the largest classes of transactions. As our results show, analysis of intermediate aggregates can provide more confidence when making audit judgments about anomalies and give the auditor a means of thinking about the underlying business process as a whole.

Summary statistics of the data used in the analysis are presented in Table 1.

[Insert Table 1 here]

As discussed above, we selected transaction item quantity as the primary metric for testing as opposed to dollar amounts, and we did so for two reasons: First, we want to illustrate that CA can work efficiently and effectively on operational (non-financial) data; second, in our sample set dollar amounts contain noisy information including sales discounts and tax. We aggregate the transaction quantities for the ordering, receiving, and voucher payment processes respectively. After excluding weekends and holidays and several observations at the beginning of the sample period to reduce noises in the sample, we have 180 days of observations in our data sets for each business process.

Continuity Equation Candidates: Simultaneous Equation Model

We investigate three probabilistic models that can serve as candidates for our continuity equation benchmarks of the firm's supply chain processes: a Simultaneous Equation Model (SEM), a Vector Autoregressive (VAR) model and a Linear Regression Model (LRM). The SEM can model the interrelationships between different business processes simultaneously while the linear regression model can only model one relationship at a time, but the latter is less computationally demanding. In SEM each interrelationship between two business processes is represented by an equation and the SEM-based CE model consists of a simultaneous system of two or more equations which represent the business processes that make up the organization.

In SEM and LRM we specify the daily aggregate of order quantity as the exogenous variable while the daily aggregates of receiving quantity and payment quantity are endogenous variables. Time stamps are added to the transaction flow among the three business processes. The transaction flow originates from the ordering process at time t. After a lag period Δ_1, the transaction flow appears in the receiving process at time t+ Δ_1. After another lag period Δ_2, the transaction flow re-appears in the voucher payment processes at time t+Δ_2. Using the SEM methodology to model these processes yields the set of equations:[15]

$$\begin{cases} (qty\ of\ \ receive)_{t+\Delta 1} \ \ = \ \ \mathrm{a}*(qty\ of\ \ order)_{t}\ +\varepsilon_1 \\ (qty\ of\ \ vouchers)_{t+\Delta 1+\Delta 2} = \ \mathrm{b}*(qty\ of\ \ receive)_{t+\Delta 1}+\varepsilon_2 \end{cases}$$

Our next step in constructing the simultaneous equation model is to estimate the lags. Initially, we used the mode and mean and other combinations of lag estimates. Our results indicate that the mode estimate works best among all estimates for the simultaneous equation model. Therefore, our estimated model is:

$$\begin{cases} receive_t \ = \ \mathrm{a}*order_{t\text{-}1}+\varepsilon_1 \\ voucher_t \ = \ \mathrm{b}*receive_{t\text{-}1}+\varepsilon_2 \end{cases}$$

---

[15] Alles et al. 2006b develop these statistical CE relationships from the underlying theoretical business processes of the firm's supply chain. In this paper adopt a purely statistical approach towards CE creation.

Where

order = daily aggregate of transaction quantity for the purchase order process

receive = daily aggregate of transaction quantity for the receiving process

voucher = daily aggregate of transaction quantity for the voucher payment process

t = transaction time

We divide our data set into two groups. The first group consisting of the first 100 days is categorized as the training set and used to estimate the model. The second group consisting of the remaining days is categorized as the hold-out set and used to test our model. Our estimated simultaneous equation model estimated on the training set is as follows:

$$\begin{cases} receive_t = 0.482 * order_{t\text{-}1} + e_1 \\ voucher_t = 0.816 * receive_{t\text{-}1} + e_2 \end{cases}$$

The R squares for the equation are 0.73 and 0.79 respectively, which indicate a good fit of data for the simultaneous equation model. However, we have also realized some limitations associated with SEM. First, the lags have to be separately estimated and such estimations are not only time-consuming but also prone to errors. Second, the SEM is a simplistic model. Each variable can only depend on a single lagged value of the other variable. For example, $voucher_t$ can only depend on $receive_{t-1}$ even though there is a strong likelihood that it can also depend on other lagged value of the *receive* variable, or even the lagged value of the *order* variable. Due to these limitations, we need to develop a more flexible CE model.

Continuity Equation Candidates: Vector Autoregressive Model

We continue to follow the BP auditing approach and use daily aggregates of transaction item quantity as audit metric to develop the VAR models. However, unlike in the case of SEM, no lag estimation is necessary. Only the maximum lag period needs to be specified. All possible lags within the period can be tested by the model. We select 13 days as the maximum lag because 90% of the lags of all the individual transactions fall within this time frame. Our basic multivariate time series model is expressed as follows:

$$order_t = \Phi_{ro} * M(receive) + \Phi_{vo} * M(voucher) + \varepsilon_o$$

$$receive_t = \Phi_{or} * M(order) + \Phi_{vr} * M(voucher) + \varepsilon_r$$

$$voucher_t = \Phi_{ov} * M(order) + \Phi_{rv} * M(receive) + \varepsilon_v$$

M (order)= n*1 vector of daily aggregate of order quantity

M (receive)= n*1 vector of daily aggregate of receive quantity

M (voucher)= n*1 vector of daily aggregate of voucher quantity

Φ = corresponding *1*n* transition vectors

Again we split our data set into two subsets: the training set and the hold-out set. SAS VARMAX procedure is used to estimate the large VAR model. Despite the fact that this model is a good fit to our data sets, the predictions it generates for the hold-out sample have large variances.[16] In addition, a large number of the parameter estimates are not statistically significant. We believe the model suffers from the over-fitting problem. Therefore, we apply step-wise procedures to restrict the insignificant parameter values to zero and retain only the significant parameters in the model in each step. Then, we estimate the model again. If new insignificant parameters appear, we restrict them to zero and re-estimate the model. We repeat the step-wise procedure several times until there are no insignificant parameters appearing in the mode, resulting in a subset VAR model. One of our estimated subset multivariate time series model is expressed as:

$$order_t = 0.24 * order_{t-4} + 0.25 * order_{t-14} + 0.56 * receive_{t-15} + e_o$$

$$receive_t = 0.26 * order_{t-4} + 0.21 * order_{t-6} + 0.60 * voucher_{t-10} + e_r$$

$$voucher_t = 0.73 * receive_{t-1} - 0.25 * order_{t-7} + 0.22 * order_{t-17} + 0.24 * receive_{t-17} + e_v$$

The over-parameterization problem can be resolved by step-wise procedures to transform the general form VAR into Subset VAR. However, it requires auditors' time and judgment to reduce the general form VAR model into Subset VAR model, which is antithetical to the automated nature of the CA system.

Recent development in Bayesian statistics, however, allows the model itself to control parameter restrictions. The BVAR model includes prior probability distribution functions to impose restrictions on the parameter estimates, with the covariance of the prior distributions controlled by "hyperparameters". In other words, the values of hyperparameters in the BVAR model control how far the model coefficients can deviate from their prior means and how much the model can approach an unrestricted VAR model (Doan et al. 1984, Felix and Nunes 2003). The BVAR model can release auditors from the burden of parameters restriction to derive the Subset VAR model.

Continuity Equation Candidates: Linear Regression Model

In the linear regression model we specify the lagged values of daily aggregates of transaction item quantity in the order process and the receive process as two independent variables respectively, and the voucher payment quantity aggregate as the dependent variable. Again, we use the mode value of lags in individual transactions as estimates for the lags in the model (i.e. 2 day lag between the ordering and voucher payment processes, and 1 day lag between the receiving and voucher payment processes). No intercept is used in our model because we can not find any valid meaning for the intercept. Our OLS linear regression model is expressed as follows:

$$voucher_t = \mathrm{a}*order_{t-2} + \mathrm{b}*receive_{t-1} + \varepsilon$$

Where

order = daily aggregate of transaction quantity for the ordering process

receive = daily aggregate of transaction quantity for the receiving process

voucher = daily aggregate of transaction quantity for the voucher payment process

t= transaction time at time t

Again we use the first 100 days of our data set as the training subset to estimate our model. The estimated linear regression model is:

$$voucher_t = 0.08*\ order_{t-2} + 0.67*\ receive_{t-1} + e$$

The a estimate is statistically insignificant ($p>0.68$) while the b estimate is significant at 99% level ($p<0.0001$).

[16] We find that the MAPEs for predictions of Order, Receive, and Voucher variables are all over 54%, much larger than the MAPEs of the Simultaneous Equation Model, the linear regression model and the subset VAR model. Refer to section 5.1 for MAPE definition.

Model Online Learning Protocol

One distinctive feature of analytical modeling in CA is the automatic model selection and updating capability. Traditional analytical modeling is usually based on static archival data sets. Auditors generally apply one model to the entire audit data set. In comparison, analytical modeling in CA can be based on the continuous data streams dynamically flowing into the CA system. The analytical modeling in CA thus has the potential to assimilate the new information contained in every segment of the data flows and adapt itself constantly. Each newly updated analytical model is used to generate a prediction only for one new segment of data. This model updating procedure is expected to improve prediction accuracy and anomaly detection capability.

[Insert Figure 3 here]

V. Anomaly Detection Comparison across CE Candidates

MAPE Comparison

While performing the analytical procedures, auditors use different methods to make predictions on account numbers. It is desirable for expectation models to make forecasts as close to actual values as possible. Many prior AP studies evaluate expectation models in terms of prediction accuracy (Kinney 1978, Wild 1987, Dzeng 1994, Allen et al. 1999, Chen and Leitch 1998, Leitch and Chen 2003). To parallel this line of research we compared the prediction accuracies for the three candidate CE models using two alternate measures of prediction accuracy, MAPE and error detecting ability.

Mean Absolute Percentage Error, the absolute value of the difference between the predicted value and the actual value measured as a percentage of the actual value, is a commonly used metric of prediction accuracy. It is expected that a good model should have a small MAPE. The training set is first used to estimate each of the four candidate models. Then, each estimated model is used to make one-step-ahead forecasts and the forecast variance calculated. After that, the model is updated based on the new data feeds in the hold-out set and the previous steps are repeated again. Finally, all the variances are summed up and divided by the total number of observations in the hold-out sample to compute the MAPE. The results for MAPE of Voucher predictions are presented in Table 2.

[Insert Table 2 Here]

The results indicate that as measured by the MAPE metric the prediction accuracies of these four models are close. The BVAR model has the best prediction accuracy (MAPE=0.3330), followed by the subset VAR model (MAPE=0.3374), though the standard deviation for the BVAR model is slightly higher than the Subset VAR. The SEM has the lowest prediction accuracy (MAPE=0.3499). These prediction accuracies indicate that the forecasts generated by the expectation models usually differ from the reported amounts by approximately 30%.

There are no universal criteria to determine whether these prediction accuracies are good or not because MAPE values are data dependent. Prior studies (Kinney 1978, Wild 1987, Chen and Leitch 1998) on expectation models indicate that large variances exist in prediction accuracies when different data sets are used. The MAPE values reported in Wild's (1987) study range from 0.012 for Cost of Goods Sold prediction to 7.6 for Cash and Security prediction using the same expectation model. Our conclusion is that by the MAPE metric, all three candidate CE models show promise as benchmarks for AP tests.

Error Detecting Ability Comparison

A continuity equation is a means towards an end and not an end in itself. The rationale for constructing a CE-based AP test is to allow the detection of anomalies effectively and efficiently. Thus while predicting the value of a variable with low MAPE is desirable, more useful is the ability to detect errors.

To measure the detection capability of the three CE candidate models we use two metrics: the number of false positive errors and the number of false negative errors.[17] A false positive error is also called a false alarm or a type I error, which is a non-anomaly mistakenly detected by the model as an anomaly. A false negative error is also called a type II error, which is an anomaly failed to be detected by the model. While a false positive error can waste auditor's time and thereby increase audit cost, a false negative error is usually more detrimental because of the material uncertainty associated with the undetected anomaly. An effective and efficient AP model should keep both the number of false positive errors and the number of false negative errors at a low level.

[17] For the presentation purpose, we also include the tables and charts of detection rate, which equals to 1 minus false negative error rate.

To compare the anomaly detection capabilities of the CE models under different settings we randomly seed eight errors into the hold-out sample. We also test how the error magnitude can affect each AP model's anomaly detection capability with five different magnitudes are used in every round of error seeding: 10%, 50%, 100%, 200% and 400% of the original actual value of the seeded observations. The entire error seeding procedure is repeated ten times to reduce selection bias and ensure randomness.[18]

Prior AP studies discuss several investigation rules to identify an anomaly (Stringer 1975, Kinney and Salaman 1982, Kinney 1987). A modified version of the statistical rule (Kinney 1987) is used in this study. Prediction intervals (PI), equivalent to a confidence interval for an individual dependent variable, are used as the acceptable threshold of variance. If the value of the prediction exceeds either the upper or lower limits of the PI, then the observation is flagged as an anomaly.

The selection of the prediction interval is a critical issue in the effectiveness of the AP test. The size of prediction interval is dictated by the value of the significance level α. Choosing a low α value (e.g. 0.01), leads to wide tolerable boundaries (i.e. large prediction interval) and a resulting low detection rate. On the other hand, if a high α value is selected, then the prediction interval would be overly narrow and many normal observations would be flagged as anomalies. To solve this problem, we have followed two approaches to select the prediction interval percentages. In the first approach α values are selected to control the number of false positive errors in various models respectively. More specifically, an α value is selected which is just large enough to yield two false positive errors in the training data set. In the second approach which is a traditional approach, predetermined α values, 0.05 and 0.1, are used for all the expectation models.

Before we can use this methodology to compare the three candidate CE models, another critical issue needs to be addressed, an issue that only arises in a continuous audit setting: real time error correction.

---

[18] Leitch and Chen (2003) use both positive and negative approaches to evaluate the anomaly detection capability of various models. In the positive approach all the observations are treated as non-anomalies. The model is used to detect those seeded errors. In contrast, the negative approach treats all observations as anomalies. The model is used to find those non-anomalies. This study only adopts the positive approach because it fits better with established audit practice for AP tests.

Real-time Error Correction

Another important distinction between CA techniques and standard auditing that was explored in this project is what we call "Real Time Error Correction". In a CA environment when an anomaly is detected, the auditor will be notified immediately and a detailed investigation will be initiated. In theory, the auditor will then have the ability to correct the error before the next round of audit starts.

Whether this technical possibility can or will be carried out in practice depends both upon the speed at which error correction can be made and the more serious issue of the potential threat to auditor independence of using data in subsequent tests that the auditor has had a role in correcting. These issues clearly require detailed consideration, but what we focused on at this stage was quantifying the benefits of real time error correction in a CA environment. These issues clearly require detailed consideration, but doing so is beyond the scope of the current study. What we focus on here is the technical implication for AP in CA if errors are indeed detected and corrected in real time in a CA environment. Specifically, when the AP model detects a seeded error in the hold-out sample, we examine the consequences on subsequent error detection if the seeded error is corrected by substitution of the original actual value before the model is used again.

For comparison purpose, we test how our candidate CE models work with and without real time error correction. Unlike continuous auditing, anomalies are detected but usually not corrected immediately in traditional auditing. To simulate this scenario, we don't correct any errors we seeded in the hold-out sample even if the AP model detects them.

[Insert Table 3A, through Table 13B here]

Overall, the results show lower false negative error rates for all three candidate CE models with error correction, especially when the error magnitude is large (100% or more). This finding is consistent with the prior expectation. Under some rare circumstances (e.g. Table 10B, Error Magnitude 50%), the non-correction model has slightly better detection rate than the correction model. All these exceptions occur when the error magnitude is small (no larger than 50 percent). Figure 4 illustrates such a scenario. The prediction interval for the next forecast value is adjusted after the first error has been detected and corrected. If the next value is an error and falls right within the prediction interval, however, won't be detected by the error-correction model. On the other hand, the error will be in the rejection area in the non-correction model.

Higher false positive error rates, mostly when error magnitudes are large, are observed for the error-correction models, which means that the error-correction models can detect more anomalies but at a cost of triggering more false alarms. However, a further investigation reveals that the false alarms are mostly caused by certain records in the holdout sample, which are named by this study as original anomalies. These original anomalies are very likely to be caused by measurement errors in the data set since our data set consists of unaudited operational data. This measurement error problem with non-audited data is also reported by previous studies (Kinney and Salamon 1982, Wheeler and Pany 1990). Because the non-correction model would not correct those undetected errors, the impact of original anomalies, whose values remain constant, would be eclipsed by the increase in seeded error magnitude. Therefore, non-correction model would trigger fewer false alarms when the magnitude in seeded error increases. On the other hand, the impact of original anomalies would not decrease as the error-correction model would correct all detected errors.

Auditors are usually more averse to false negative errors than to false positive errors if the false positive error rate is kept at a reasonable level. The cost of false positive errors is only a waste of auditor's time and effort while the cost of false negative errors can be detrimental to the client firm and auditor's reputation. In summary, the error-correction models have better anomaly detection performance than the non-correction models. The error correction protocol can improve the anomaly detection performance.

Another noteworthy finding is that the α value controls the tradeoff between the false positive and false negative error rates. A large α value leads to more false positive errors but fewer false negative errors. Meanwhile, a small α value leads to fewer false positive errors but more false negative errors. It should also be noted that even though the BVAR model generally has the best detection performance, its false positive rate is also the highest among all models. Another finding is that even though we use the α value which only yields two false positive errors in the training data set for each CE model, the number of false positive errors generated in the hold-out sample is not equal among the CE models. This study does not control the number of false positive errors in the hold-out sample because it can lead to a look-ahead problem. Specifically, the data in the hold-out sample can not be used to construct CE models since they would not be available in the real world when the auditor builds expectation models.

Disaggregated data versus aggregated data

It is expected that unconstrained data are not only available but also used by auditors in the proposed CA system, which leaves the choice of the degree of aggregation in the AP tests up to them. Prior studies (Kinney and Salamon 1982, Wheeler and Pany 1990, Dzeng 1994) suggest that disaggregated data can provide better performance for expectation models. This study examines if the disaggregated data can make CE models perform better than the aggregated data. Data can be aggregated on different dimensions, and we compare the efficacy of CE based AP tests using temporal and geographic disaggregation.

In the temporal disaggregation analysis we examine the differential anomaly detection performances using weekly data versus the daily data. Errors are seeded into the weekly data in the same fashion as in previous simulations. We follow prior studies (Kinney and Salamon 1982, Wheeler and Pany 1990) to seed errors into the daily data. In the best case scenario, the entire weekly error is seeded into a randomly selected day of a week. In the worst case scenario, the weekly error is first divided by the number of working days in a week (e.g. E/5) and then seeded into each working day of that week. In addition to the different aggregation levels of data comparison, the error-correction and non-correction models are again compared to verify if the previous findings still hold. Due to the scope of this study we only use a single α value 0.05 in all models. The results are presented in Tables 14A/B through Tables 17A/B and Figures 4A/B through Figures 7A/B.

The results are generally consistent with our expectations. In terms of detection ability, all the CE models perform the best using the best case scenario daily data, followed by the weekly data. All the models have the poorest anomaly detection performance using the worst case scenario daily data. This result is not surprising because the weekly error is spread evenly into each day making the seeded error act as a systematic error which is almost impossible to detect (Kinney 1978). With respect to false positive error rates, the results are mixed. We believe that the original anomalies in our data sets caused this problem. In the cross model comparison the BVAR model generally detects more errors than other models but in the mean time triggers more false alarms. The linear regression model generally has fewer false alarms but suffers from low detection rate when error magnitudes are small.

We repeat the analyses aggregating data on the geographic dimension and obtain similar results.

VI Discussion

Limitations of the Analysis

Our data sets are extracted from a single firm, which may constitute a selection bias. Until we test our data level CA system using other firms' data sets, we will not have empirical evidence to support that our AP models are portable and can be applied to other firms. In addition, our data sets contain noise, as the fact that there are preexisting anomalies in our data indicates. Since our data sets are actually extracted from a central data warehouse which accepts data from both ERP and legacy systems in the firm's subdivisions, it is inevitable for our data sets to be contaminated by errors and noises. The date truncation problem also produces noise in our data sets. Of course, all AP tests suffer from these problems, and they are not unique to the CA environment.

As with any analytical procedures, a detected anomaly can only indicate the presence of a problem, and cannot pinpoint the problem itself, while a failed test of detail (for example, a negative confirmation, a reconciliation failure) does, but only if the auditor knows which data to test. Business processes can break down for a variety of reasons, some "real", meaning at the business process level itself, and some "nominal", meaning that even if the integrity of the underlying business process is not compromised, the CE may fail to represent that.

An example of a "nominal" violation would be a seasonal slow down in the delivery of shipments, which results in a broken CE model due to a shift in the value of the time lag. This is not indicative of a faulty business process, but an inevitable outcome of trying to fit the changing reality into a benchmark constructed using obsolete data. Thus, the auditor's investigation is bound to identify this situation as a false positive, unless the CE model is able to adapt accordingly.

The CE model is expected to signal the presence of anomalies in cases where the underlying business process is compromised, as for example when a strike affects a supplier or when a raw material becomes scarce. The purpose of using CE-based AP tests is to detect these process errors and to then generate a signal for the auditor to investigate the reasons for the broken processes through a targeted investigation of details in as real time as possible. This clearly shows the advantage of using continuity equations to relate the most disaggregated metrics possible, since the more disaggregated the metrics are the narrower the scope of the auditor's investigation can be. However, the more disaggregated the metrics, the less stable the CE relationship. This is the inescapable tradeoff between the level of disaggregation of the metrics and the stability of the continuity equations, as the results of this study demonstrate.

Another issue emerged in this study is the selection of the α value which controls the number of false positive errors and false negative errors. The optimal α value cannot be obtained unless the costs of false positive errors and false negative errors are known. However, while it is generally accepted that the cost of false negative errors greatly exceeds that of false positive errors, the choice of a particular ratio of these costs is usually highly controversial. Additionally, the optimal α value would change as data sets and expectation models change.

Future Research Directions

Since this paper is devoted to a new research area, much work needs to be done. The most important issue that has to be addressed is the feasibility of using CE models in practice. Few auditors in the field will be willing or able to use statistical methods as demanding as those utilized in this project. For continuous data level auditing to be truly useful in future CA systems, there will have to be close to a "push button" addition to the audit toolkit, which means that at a minimum CE models developed in the laboratory must be generally applicable to different firms and processes. Testing the robustness of the CE models created using this data on other data sets is on our research agenda. CE models should also be compared with other approaches to BP modeling such as artificial intelligence. The value of including other independent and control variables in our models also needs to be examined.

An important step when implementing continuous data level auditing is in determining which business process should be modeled by continuity equations and subject to AP testing. There are a large number of business processes that can be modeled—and as we have seen, far more data than can be handled—and the auditor cannot model all of them and indeed, does not need to. Providing assurance only necessitates the examination of the key processes that define a business, and these, in turn, will depend on the company's strategy, its product space, competitive environment and current financial condition and history. There needs to be research undertaken on the general question of the types of business processes that an auditor is likely to encounter, and the way in which these processes can be represented in continuity equations.

In particular, it needs to be kept in mind that while certain important indicators of business processes can be implicit and conceptual (for example, "customer satisfaction is important because it ultimately drives profits"), continuity equations are meant to be used in tests of factual evidence. That measurement aspect of continuity equations is why they are not synonymous with a business process, but only a measurable analytical model of one. That is, the continuity equation relates the metrics, while the business process is the underlying set of activities, and so the continuity equations model is a function not only of the characteristics of that business process, but of the way in which that process is measured. That fact also means that as a first step towards a categorization of continuity equations and their characteristics, we have to develop a classification of business processes:

Endogenous or Exogenous: Endogenous processes are ones determined by underlying technological or institutional constraints, and so are under the control of the company, while exogenous processes are ones whose outcomes depend on the actions of other parties. Outsourcing relationships and special purpose entities, however, can obscure clarity of this classification. The importance of supply chain management and globalization emphasizes the need for auditors to take a broad view of the scope of the businesses, looking beyond functions, departments and even the company itself. Thus the value chain is extended upstream, to important suppliers (who are increasingly, global), and downstream, to customers.

Degree of Uncertainty: Closely related to the prior classification is the degree of uncertainty of the process, which is defined over a continuum because there are very few processes that are entirely certain. Even such endogenous processes as production functions will have normal variation in output, while exogenous processes are hypothesized relationships to begin with, and so have inherent uncertainty. Business process uncertainty will carry through to the continuity equation built upon it, and the nature of that process will also affect the level of measurement uncertainty of the continuity equation itself.

Financial or Non Financial: Continuity equations based on the accounting relationships, such as reconciliations, are financial in nature, by definition. While an important task in the audit is establishing the reliability of accounting numbers (especially given the requirements of Section 404 of the Sarbanes-Oxley Act), there is now a widespread recognition that managing the company requires looking beyond the financial numbers to the underlying non-financial variables and processes. Thus, the Total Quality Control movement pointed out that quality only improves when it is measured directly, on a per-unit basis, rather than at the aggregated cost of quality level. The challenge facing auditors today is incorporating the analysis of non-financial drivers of company performance into their examination of financial measures of profit.

The fact that continuity equations provide an analytical model of interrelated business processes means that they also can be classified according to the categories listed above. But the CE model is not synonymous with the underlying business processes. Even if the underlying business process is conceptually certain, the continuity equation representation of it could be uncertain because the measurement of the data takes place within bounds of precision and time, thus a distinction that must be made between the underlying business process and the probabilistic CE model that is used to provide a benchmark for it.

Statistical and non-statistical Methods: The CE models used in the analytical procedures are based on statistical methods. One can introduce non-statistical methods to future studies and compare the results with the CE models. Artificial neural networks can be a good option since their application in forecasting has been praised by other researchers.

Much more work obviously needs to be done in developing and implementing CE models, combining theoretical analysis with empirical research that will illustrate the challenges that auditors will encounter when putting the concept of continuous data level auditing into practice. In particular, the costs and benefits of CE based analytic procedures need further investigation.

VII Conclusion

In this paper we develop a continuous data level auditing system utilizing business process based analytic procedures. The first component of the system is automatic transaction verification to filter out exceptions, which are transactions violating user defined business process rules. The second component of the CA system uses continuity equations to provide benchmarks for process level AP tests which are applied to the transaction stream to identify statistical anomalies possibly indicating business process problems. AP tests are needed to complement transaction verification given the inability to define rules for all possible exceptions and to identify patterns of anomalous firm behavior.

This is the first study on the use of unconstrained data to develop analytical procedures for continuous auditing. It is also the first attempt to use empirical data to compare different AP models in a CA context. The choice of aggregation level in the AP tests is up to the auditor who is not constrained to use data that is already at a high level of aggregation, such as account balances. Hence, the auditor has the freedom to make the tradeoff between utilizing more stable metrics by appropriate aggregation versus the resulting loss of information content, as opposed to being forced to accept limitation on the data imposed by outside circumstances. The highly disaggregated data that underlies CA allows auditors to fit equations relating to specific metrics, such as those related to individual business units, or individual customers or vendors, or small coherent subgroups of them. The main benefit, as compared with traditional analytical procedures is that anomalies can be identified which are undetectable at higher levels of aggregation.

We model flow relationships between different metrics of related business processes as a system of continuity equations. We use a seeded error simulation study to compare the anomaly detection capability of three candidate CE models. Our results confirm that joint analysis of business processes gives the auditor an analytical procedure with a robust capability to detect anomalies in a real time continuous auditing environment with highly disaggregated data.

An important methodological innovation of this study is the examination of the capabilities of CA to investigate and correct identified problems in (close to) real time. We therefore introduce a real time error correction protocol in our simulation study and examine the differential detection capabilities between models with error correction and without error correction. We show that, as expected, under most circumstances the use of real time error correction results in superior performance.

Overall this study finds that while there are differences in the predictive ability and anomaly detection performance of candidate CE models, all models perform extremely well and no single model performs better on all aspects. From this we can draw two important conclusions.

First, that unlike in the traditional audit literature on analytic procedures, the inability to pick a winner in a "horse race" across the candidate CE models is less important than the fact that all models yield highly efficient AP tests. The point is that because of its automated and technology driven nature, it is quite feasible and even desirable for the continuous data level audit system to use benchmarks based on multiple CE models instead of being forced to select only one, as would be necessary in a more manual system. For example, the BVAR model can be used first to detect anomalies because it has a low false negative error rate. Subsequently, the simultaneous equation model and the linear regression model can be used to remove the false alarms from the BVAR-detected anomalies because these two models have relatively low false positive error rates.

The point to be remembered is that in a CA setting with unconstrained data combined with unconstrained computational resources to analyze that data there is no longer any need to be parsimonious in the collection of audit evidence, yet another indication of the different mentality that auditors will need to adopt as they shift from standard to continuous auditing, Indeed, dealing with more data rather than less, in real time as opposed to archival, will become a necessity once stakeholders recognize that traditional audit methodologies are essentially throwing away the richness of the unconstrained data that the firm's IT systems are now making available to the auditor, not to mention the fact that the time frame for the audit is increasingly at odds with the decision cycles of the real time business, as well as its process driven strategy.

Our second conclusion from the fact that all three CE models yield effective analytic procedures is that when auditors have access to unconstrained data, the richness of that disaggregate data combined with the ability to make real time error correction makes error detection robust across a variety of expectations models. In other words, it is the nature of the data that serves as audit evidence that is the primary driver of audit effectiveness, with the selection of the specific analytic procedure a second order concern—not because the audit benchmark is not important, but because auditing at the process level makes errors stand out much more obviously in the data.

Of course, we obtain this result in a setting in which our AP tests use continuity equation based benchmarks that are already sophisticated models of underlying business processes. Thus perhaps the correct way of stating our conclusion is that when data is unconstrained, it is important to not throw away that data and to conduct analytic procedures at the process level. Doing so necessitates the use of continuity equations as benchmarks, but the auditor has the reassurance when implementing data level CA that any CE model will work almost as well.

These are key conclusions in support continuous data level auditing and its ability to make use of unconstrained data, implement real time error correction and give auditors the choice of the degree of aggregation. Future research applying this CA methodology across different firms and different data sets is necessary to see whether these conclusions are robust, and to examine what their implications are for audit practice.

References:

1. Allen R.D., M.S. Beasley, and B.C. Branson. 1999. Improving Analytical Procedures: A Case of Using Disaggregate Multilocation Data, *Auditing: A Journal of Practice and Theory* 18 (Fall): 128-142.
2. Alles M.G., G. Brennan A. Kogan, and M.A. Vasarhelyi. 2006. Continuous Monitoring of Business Process Controls: A Pilot Implementation of a Continuous Auditing System at Siemens. *International Journal of Accounting Information Systems*, Volume 7, Issue 2 (June): 137-161.
3. Alles M.G., A. Kogan, and M.A Vasarhelyi. 2006b. Continuous Auditing: Lessons from Fifteen Years of Transforming Theory into Practice. Working paper, Rutgers Business School.
4. ____. 2002. Feasibility and Economics of Continuous Assurance. *Auditing: A Journal of Practice and Theory* 21 (Spring):125-138.
5. ____.2004. Restoring auditor credibility: tertiary monitoring and logging of continuous assurance systems. *International Journal of Accounting Information Systems* 5: 183-202.
6. ____ and J. Wu. 2004. Continuity Equations: Business Process Based Audit Benchmarks in Continuous Auditing. *Proceedings of American Accounting Association Annual Conference.* Orlando, FL.
7. American Institute of Certified Public Accountants. 1988. Statement on Auditing Standards No. 56: Analytical Procedures. New York.
8. Auditing Concepts Committee of the American Accounting Association. 1972. Report of the committee on basic auditing concepts.
9. Bell T., Marrs F.O., I. Solomon, and H. Thomas 1997. *Monograph: Auditing Organizations Through a Strategic-Systems Lens*. Montvale, NJ, KPMG Peat Marwick.
10. Biggs, S., T. Mock, and P Watkins. 1988. Auditors' use of analytical review in audit program design. *The Accounting Review* 63 (January): 148-161
11. Brown, Carol E. Jeffrey A. Wong, and Amelia A. Baldwin. 2006. Research Streams in Continuous Audit: A Review and Analysis of the Existing Literature. *Collected Papers of the Fifteenth Annual Research Workshop on: Artificial Intelligence and Emerging Technologies in Accounting, Auditing and Tax*. pp. 123-135. Washington, DC, USA, August 5, 2006.

12. Chen Y. and Leitch R.A. 1998. The Error Detection of Structural Analytical Procedures: A Simulation Study. *Auditing: A Journal of Practice and Theory* 17 (Fall): 36-70.
13. _______________________. 1999. An Analysis of the Relative Power Characteristics of Analytical Procedures. *Auditing: A Journal of Practice and Theory* 18 (Fall): 35-69.
14. CICA/AICPA. 1999. *Continuous Auditing.* Research Report, Toronto, Canada: The Canadian Institute of Chartered Accountants.
15. Davenport, T.H. and J.E. Short, 1990. The New Industrial Engineering: Information Technology and Business Process Redesign, *Sloan Management Review*, pp. 11-27, Summer.
16. Doan, T.A., Littleman R.B., and C.A. Sims. 1984. Forecasting and Conditional Projections Using Realistic Prior Distributions. *Econometric Reviews* 1, 1-100.
17. Dzeng S.C. 1994. A Comparison of Analytical Procedures Expectation Models Using Both Aggregate and Disaggregate Data. *Auditing: A Journal of Practice and Theory* 13 (Fall), 1-24.
18. Elliot, R.K. 2002. Twenty-First Century Assurance. *Auditing: A Journal of Practice and Theory*. 21 (Spring), 129-146.
19. Felix, R.M. and L.C. Nunes. 2003. Forecasting Euro Area Aggregates with Bayesian VAR and VECM Models. Working Paper. Banco De Portugal. Economic Research Department.
20. Geerts, G. and W. McCarthy. 1997. Modeling Business Enterprises as Value-Added Process Hierarchies with Resource-Event- Agent Object Templates, in *Business Object Design and Implementation* J. Sutherland and D. Patel (eds.), 1997, Springer-Verlag, 94-113.
21. _______________________. 2005. Working paper. Using Object Templates from the REA Accounting Model to Engineer Business Processes and Tasks.
22. Groomer, S.M. and U.S. Murthy. 1989. Continuous auditing of database applications: An embedded audit module approach. *Journal of Information Systems 3* (2), 53-69.
1. Hammer, M. 1990. Reengineering Work: Don't Automate, Obliterate! *Harvard Business Review.*

23. Hirst, E. and L. Koonce. 1996. Audit Analytical Procedures: A Field Investigation.*Contemporary Accounting Research.*, Vol 13, No 2, Fall.

24. Hylas, R. and R. Ashton. 1982. Audit detection of financial statement errors, *The Accounting Review*, Vol. 57 No.4, 751-65.

25. Kinney, W.R. 1978. ARIMA and Regression in Analytical Review: an Empirical Test. *The Accounting Review*, Vol. 17, No. 1, 148-165.

26. Kinney, W.R. and G.L. Salamon. 1982. Regression Analysis in Auditing: A Comparison of Alternative Investigation Rules. *Journal of Accounting Research*. Vol. 20. No. 2, 350-366.

27. Kinney, W.R.1987. Attention-Directing Analytical Review Using Accounting Ratios: A Case Study. *Auditing: A Journal of Practice and Theory* Vol. 6 No.2 (Spring), 59-73.

28. Knechel, W. R. 1988. The effectiveness of statistical analytical review as a substantive auditing procedure: A simulation analysis. *The Accounting Review* (January), 74-95.

29. Kogan, A. E.F. Sudit, and M.A. Vasarhelyi. 1999. Continuous Online Auditing: A Program of Research. *Journal of Information Systems* 13 (Fall), 87–103.

30. Koreisha, S. and Y. Fang. 2004. Updating ARMA Predictions for Temporal Aggregates. *Journal of Forecasting* 23, 275-396.

31. Leitch and Y. Chen. 2003. The Effectiveness of Expectation Models In Recognizing Error Patterns and Eliminating Hypotheses While Conducting Analytical Procedures. *Auditing: A Journal of Practice and Theory* 22 (Fall), 147-206.

32. Loebbecke, J. and P., Steinbart. 1987. An investigation of the use of preliminary analytical review to provide substantive audit evidence, *Auditing: A Journal of Practice and Theory*, Vol. 6 No.2, 74-89.

33. Lorek, K.S., B.C. Branson, and R.C. Icerman. 1992. On the use of time-series models as analytical procedures. *Auditing: A Journal of Practice & Theory*, Vol. 11, No. 2 (Fall), 66-88.

34. McCarthy, W. 1979. An Entity-Relationship View of Accounting Models, *The Accounting Review*, 667-86.

35. ___________. 1982. The REA Accounting Model: A Generalized Framework for Accounting Systems in a Shared Data Environment. *The Accounting Review,* 554-78.

36. Murthy, U.S. 2004. An Analysis of the Effects of Continuous Monitoring Controls on e-Commerce System Performance. *Journal of Information Systems*. 18 (Fall), 29–47.

37. ____________and M.S. Groomer. 2004. A continuous auditing web services model for XML-based accounting systems. *International Journal of Accounting Information Systems* 5, 139-163.
38. Pandher G.S. 2002. Forecasting Multivariate Time Series with Linear Restrictions Using Unconstrained Structural State-space Models. *Journal of Forecasting* 21: 281-300.
39. Porter, M. E. 1996. What Is Strategy? *Harvard Business Review*, Vol. 74, #6, S. 61-78.
40. PricewaterhouseCoopers. 2006. State of the internal audit profession study: Continuous auditing gains momentum. http://www.pwcglobal.com/images/gx/eng/about/svcs/grms/06_IAState_Profession_Study.pdf
41. Rezaee, Z., A. Sharbatoghlie, R. Elam, and P.L. McMickle. 2002. Continuous Auditing: Building Automated Auditing Capability. *Auditing: A Journal of Practice and Theory* 21 (Spring), 147-163.
42. Searcy, D. L., Woodroof, J. B., and Behn, B. 2003. Continuous Audit: The Motivations, Benefits, Problems, and Challenges Identified by Partners of a Big 4 Accounting Firm. *Proceedings of the 36th Hawaii International Conference on System Sciences*: 1-10.
43. Stringer, K. and T. Stewart. 1986. *Statistical techniques for analytical review in auditing*. Wiley Publishing. New York.
44. Swanson, N., E. Ghysels, and M. Callan. 1999. A Multivariate Time Series Analysis of the Data Revision Process for Industrial Production and the Composite Leading Indicator. Book chapter of *Cointegration, Causality, and Forecasting: Festchrift in Honour of Clive W.J. Granger*. Eds. R. Engle and H. White. Oxford: Oxford University Press.
45. Vasarhelyi, M.A and F.B. Halper. 1991. The Continuous Audit of Online Systems. *Auditing: A Journal of Practice and Theory* 10 (Spring), 110–125.
46. ________________ 2002. Concepts in Continuous Assurance. Chapter 5 in *Researching Accounting as an Information Systems Discipline*, Edited by S. Sutton and V. Arnold. Sarasota, FL: AAA.
47. _______________, M. Alles, and A. Kogan. 2004. Principles of Analytic Monitoring for Continuous Assurance. *Journal of Emerging Technologies in Accounting*, Vol. 1, 1-21.
48. _______________, and M. Greenstein. 2003. Underlying principles of the electronization of business: A research agenda. *International Journal of Accounting Information Systems* 4: 1-25.

49. Wright, A. and R.H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. The Accounting Review (October), 710-28.

50. Woodroof, J. and D. Searcy 2001. Continuous Audit Implications of Internet Technology: Triggering Agents over the Web in the Domain of Debt Covenant Compliance. *Proceedings of the 34th Hawaii International Conference on System Sciences*.

Figures, Tables and Charts

Figure 1: Architecture of Data-Oriented Continuous Auditing System

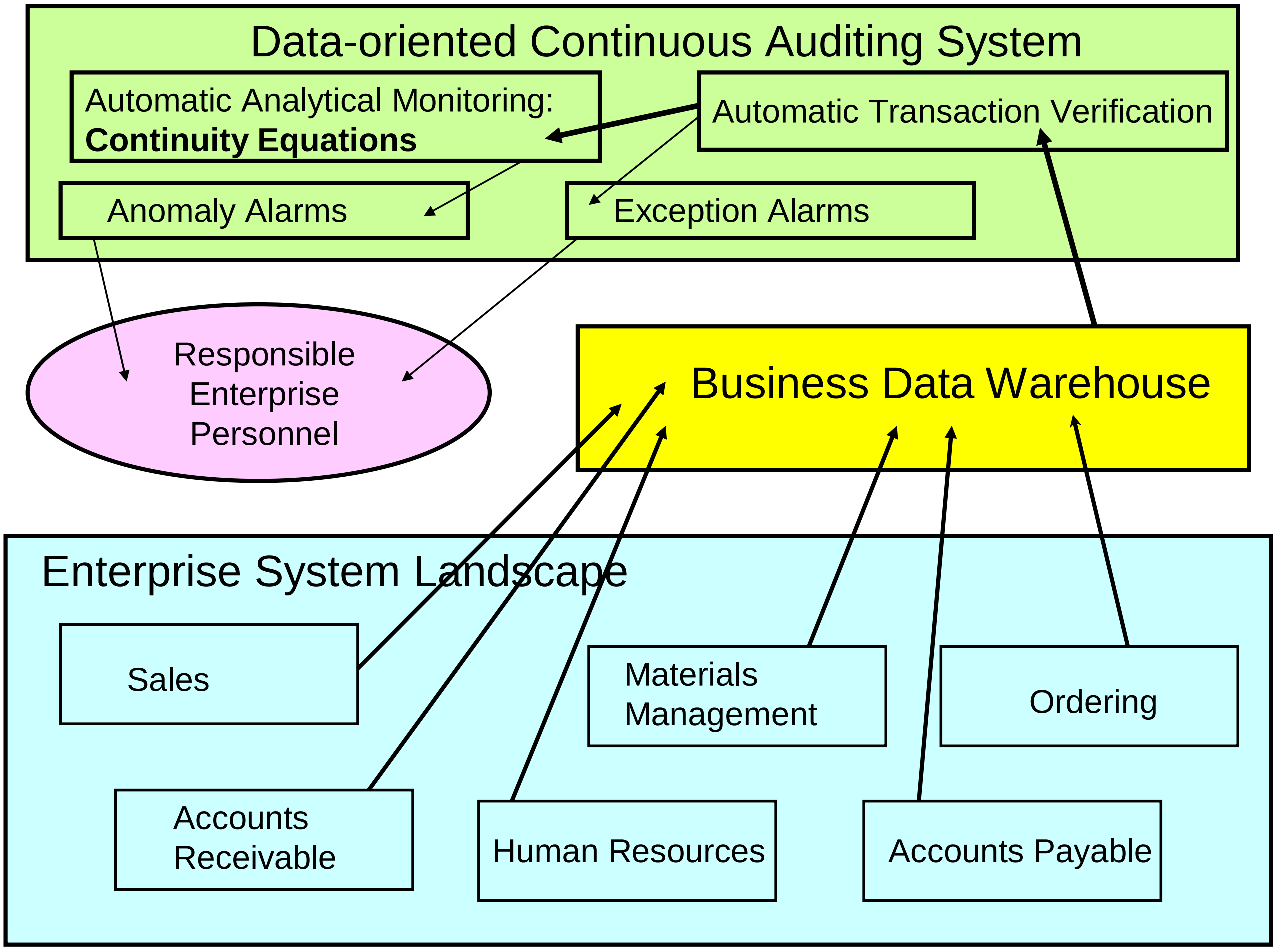


Figure 2: Business Process Transaction Flow Diagram

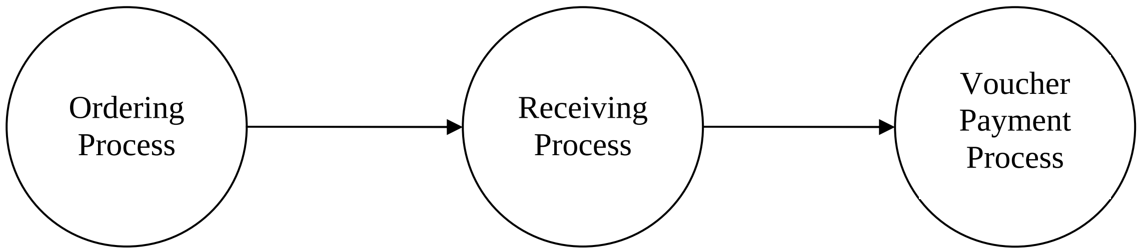


Figure 3: Model Updating Protocol

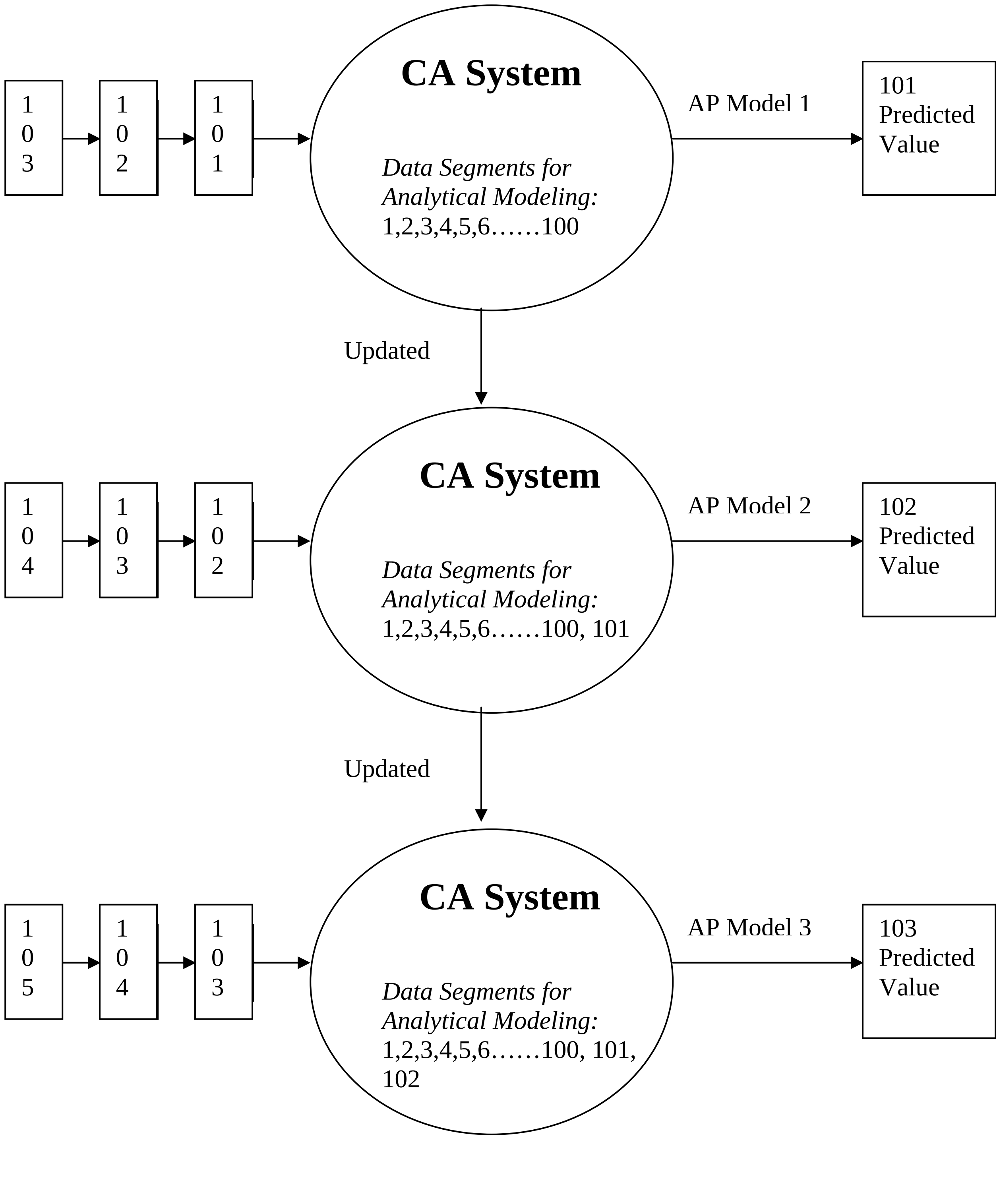


Figure 3: Multivariate Time Series Model Selection

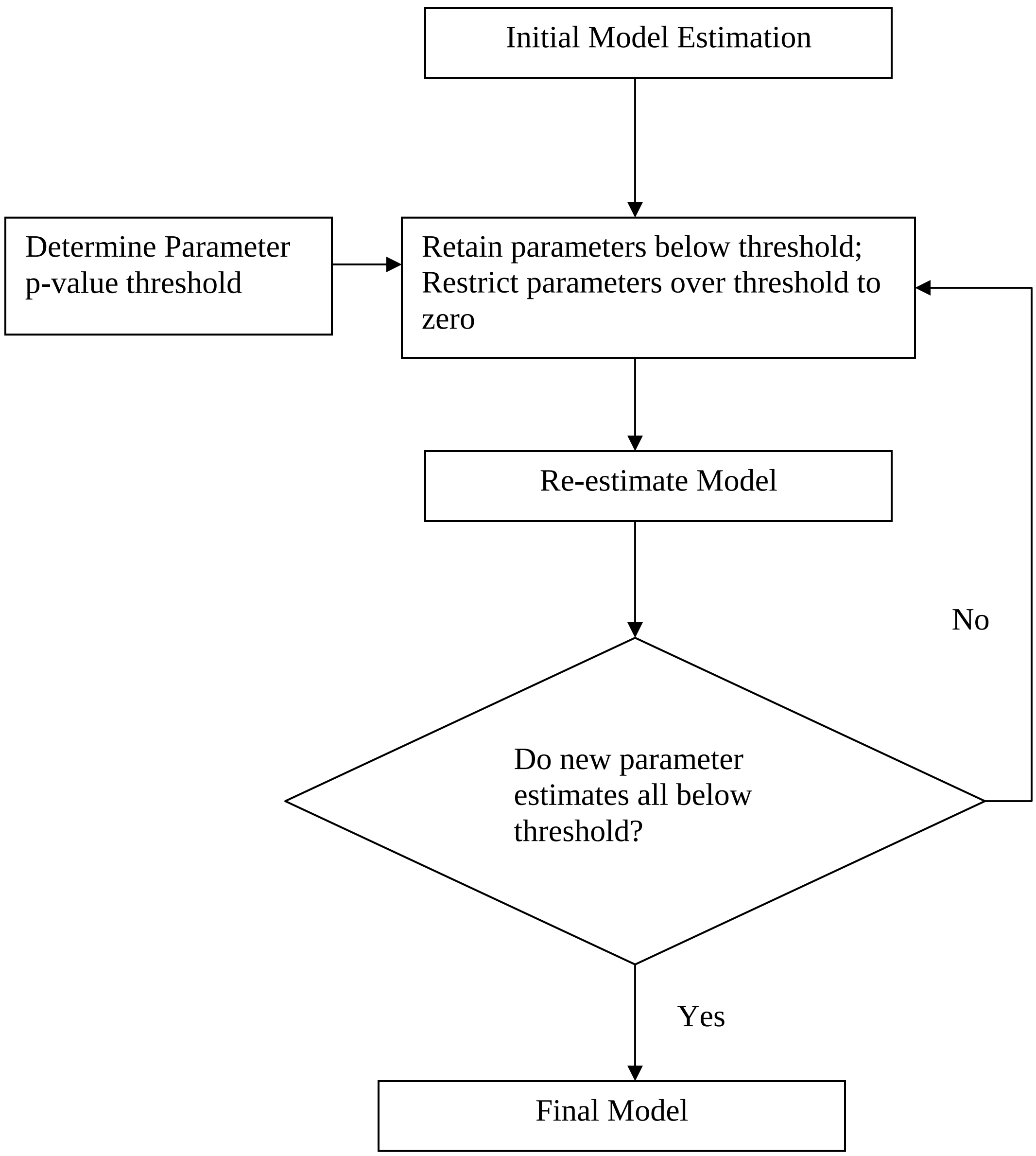


Table 1: Summary Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
Order	180	10631.38	6840.51	4597	51392
Receive	180	8876.26	3195.16	4430	31412
Voucher	180	7496.74	3874.47	187	31209

The table presents the summary statistics for the transaction quantity daily aggregates for each business process. The low minimums for Receive and Voucher are due to the date cutting off problem. The data sets span from 10/01/03 to 06/30/04. Many related transactions for the Receive and Voucher for the first 2 days in the data sets may happen before 10/01/03.

Table 2A: False Positive Error Rates Comparison between Error-correction and Non-correction Bayesian MTSM Using Daily Quantity Aggregate Data of Entire Company, α=0.18

Error Magnitude	Error Correction	Non-Correction
10%	0.4597	0.4597
50%	0.4597	0.4472
100%	0.4597	0.4486
200%	0.4597	0.4125
400%	0.4597	0.3722

Table 2B: False Negative Error Rates Comparison between Error-correction and Non-correction Bayesian MTSM Using Daily Quantity Aggregate Data of Entire Company, α =0.18

Error Magnitude	Error Correction	Non-Correction
10%	0.4625	0.4625
50%	0.25	0.25
100%	0.1125	0.125
200%	0.05	0.05
400%	0	0

Table 3A: False Positive Error Rates Comparison between Error-correction and Non-correction Bayesian MTSM Using Daily Quantity Aggregate Data of Entire Company, $\alpha = 0.10$

Error Magnitude	Error Correction	Non-Correction
10%	0.3847	0.3847
50%	0.3833	0.3806
100%	0.3833	0.3681
200%	0.3833	0.3542
400%	0.3833	0.2958

Table 3B: False Negative Error Rates Comparison between Error-correction and Non-correction Bayesian MTSM Using Daily Quantity Aggregate Data of Entire Company, $\alpha = 0.10$

Error Magnitude	Error-Correction	Non-Correction
10%	0.5152	0.5125
50%	0.2875	0.3250
100%	0.1625	0.1875
200%	0.0625	0.1000
400%	0	0.025

Table 4A: False Positive Error Rates Comparison between Error-correction and Non-correction Bayesian MTSM Using Daily Quantity Aggregate Data of Entire Company, $\alpha = 0.05$

Error Magnitude	Error Correction	Non-Correction
10%	0.342	0.336
50%	0.339	0.328
100%	0.340	0.318
200%	0.344	0.293
400%	0.344	0.250

Table 4B: False Negative Error Rates Comparison between Error-correction and Non-correction Bayesian MTSM Using Daily Quantity Aggregate Data of Entire Company, $\alpha = 0.05$

Error Magnitude	Error-Correction	Non-Correction
10%	0.5500	0.5500
50%	0.3750	0.4000
100%	0.2125	0.2500
200%	0.0625	0.1250
400%	0	0.0250

Table 5A: False Positive Error Rates Comparison between Error-correction and Non-correction Subset MTSM Models Using Daily Quantity Aggregate Data of Entire Company, α =0.15

Error Magnitude	Error Correction	Non-Correction
10%	0.1083	0.1083
50%	0.1083	0.1000
100%	0.1083	0.0875
200%	0.1083	0.0292
400%	0.1083	0.0014

Table 5B: False Negative Error Rates Comparison between Error-correction and Non-correction Subset MTSM Models Using Daily Quantity Aggregate Data of Entire Company, α =0.15

Error Magnitude	Error Correction	Non-Correction
10%	0.8125	0.8125
50%	0.7250	0.7125
100%	0.3750	0.4625
200%	0.1000	0.2250
400%	0	0.1125

Table 6A: False Positive Error Rates Comparison between Error-correction and Non-correction Subset MTSM Models Using Daily Quantity Aggregate Data of Entire Company, α =0.10

Error Magnitude	Error Correction	Non-Correction
10%	0.088	0.083
50%	0.082	0.074
100%	0.088	0.064
200%	0.088	0.032
400%	0.088	0.019

Table 6B: False Negative Error Rates Comparison between Error-correction and Non-correction Subset MTSM Models Using Daily Quantity Aggregate Data of Entire Company, α =0.10

Error Magnitude	Error Correction	Non-Correction
10%	0.8625	0.8625
50%	0.8000	0.8000
100%	0.5000	0.6000
200%	0.1250	0.2125
400%	0	0.1250

Table 7A: False Positive Error Rates Comparison between Error-correction and Non-correction Subset MTSM Models Using Daily Quantity Aggregate Data of Entire Company, α =0.05

Error Magnitude	Error Correction	Non-Correction
10%	0.039	0.039
50%	0.039	0.025
100%	0.039	0.014
200%	0.039	0.004
400%	0.039	0.002

Table 7B: False Negative Error Rates Comparison between Error-correction and Non-correction Subset MTSM Models Using Daily Quantity Aggregate Data of Entire Company, α =0.05

Error Magnitude	Error Correction	Non-Correction
10%	0.925	0.925
50%	0.825	0.825
100%	0.615	0.663
200%	0.212	0.350
400%	0.013	0.200

Table 8A: False Positive Error Rates Comparison between Error-correction and Non-correction Simultaneous Equation Models Using Daily Quantity Aggregate Data of Entire Company, α =0.04

Error Magnitude	Error Correction	Non-Correction
10%	0.1000	0.1000
50%	0.1014	0.1029
100%	0.1043	0.0800
200%	0.1029	0.0556
400%	0.0986	0.0271

Table 8B: False Negative Error Rates Comparison between Error-correction and Non-correction Simultaneous Equation Models Using Daily Quantity Aggregate Data of Entire Company, α =0.04

Error Magnitude	Error Correction	Non-Correction
10%	0.8375	0.8375
50%	0.7750	0.7750
100%	0.4375	0.4750
200%	0.2000	0.2625
400%	0.0250	0.1625

Table 9A: False Positive Error Rates Comparison between Error-correction and Non-correction Simultaneous Equation Models Using Daily Quantity Aggregate Data of Entire Company, α =0.10

Error Magnitude	Error Correction	Non-Correction
10%	0.1886	0.1886
50%	0.1900	0.1886
100%	0.1912	0.1757
200%	0.1887	0.1257
400%	0.1864	0.0629

Table 9B: False Negative Error Rates Comparison between Error-correction and Non-correction Simultaneous Equation Models Using Daily Quantity Aggregate Data of Entire Company, α =0.10

Error Magnitude	Error Correction	Non-Correction
10%	0.75	0.75
50%	0.687	0.7
100%	0.337	0.375
200%	0.137	0.2
400%	0.012	0.1

Table 10A: False Positive Error Rates Comparison between Error-correction and Non-correction Simultaneous Equation Models Using Daily Quantity Aggregate Data of Entire Company, α =0.05

Error Magnitude	Error Correction	Non-Correction
10%	0.1386	0.1386
50%	0.1342	0.1200
100%	0.1412	0.1000
200%	0.1400	0.0629
400%	0.1331	0.0286

Table 10B: False Negative Error Rates Comparison between Error-correction and Non-correction Simultaneous Equation Models Using Daily Quantity Aggregate Data of Entire Company, α =0.05

Error Magnitude	Error Correction	Non-Correction
10%	0.8125	0.8125
50%	0.775	0.775
100%	0.425	0.425
200%	0.1875	0.25
400%	0.025	0.1125

Table 11A: False Positive Error Rates Comparison between Error-correction and Non-correction Linear Regression Models Using Daily Quantity Aggregate Data of Entire Company, α =0.06

Error Magnitude	Error Correction	Non-Correction
10%	0.1000	0.0971
50%	0.0900	0.0886
100%	0.0900	0.0571
200%	0.9710	0.0343
400%	0.1000	0.0200

Table 11B: False Negative Error Rates Comparison between Error-correction and Non-correction Linear Regression Models Using Daily Quantity Aggregate Data of Entire Company, α =0.06

Error Magnitude	Error Correction	Non-Correction
10%	0.8500	0.8500
50%	0.8125	0.8125
100%	0.4875	0.5750
200%	0.1750	0.2875
400%	0.0125	0.1500

Table 12A: False Positive Error Rates Comparison between Error-correction and Non-correction Linear Regression Models Using Daily Quantity Aggregate Data of Entire Company, α =0.10

Error Magnitude	Error Correction	Non-Correction
10%	0.1114	0.1071
50%	0.1	0.1
100%	0.1043	0.0943
200%	0.1086	0.0486
400%	0.11	0.0271

Table 12B: False Negative Error Rates Comparison between Error-correction and Non-correction Linear Regression Models Using Daily Quantity Aggregate Data of Entire Company, α =0.10

Error Magnitude	Error Correction	Non-Correction
10%	0.8125	0.8125
50%	0.7625	0.7625
100%	0.4125	0.4625
200%	0.1375	0.225
400%	0.0125	0.0875

Table 13A: False Positive Error Rates Comparison between Error-correction and Non-correction Linear Regression Models Using Daily Quantity Aggregate Data of Entire Company, α =0.05

Error Magnitude	Error Correction	Non-Correction
10%	0.09	0.09
50%	0.089	0.0643
100%	0.089	0.0457
200%	0.09	0.0286
400%	0.09	0.02

Table 13B: False Negative Error Rates Comparison between Error-correction and Non-correction Linear Regression Models Using Daily Quantity Aggregate Data of Entire Company, α=0.05

Error Magnitude	Error Correction	Non-Correction
10%	0.8625	0.8625
50%	0.8125	0.8125
100%	0.525	0.6125
200%	0.1875	0.2875
400%	0.0125	0.175

Table 14A: BVAR False Positive Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company	Mean Subunit Best Case	Mean Subunit Worst Case

	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.342	0.336	0.344	0.336	0.347	0.347
50%	0.339	0.328	0.350	0.275	0.347	0.339
100%	0.340	0.318	0.347	0.244	0.344	0.336
200%	0.344	0.293	0.349	0.210	0.344	0.320
400%	0.344	0.250	0.349	0.173	0.350	0.285

Table 14B: BVAR False Negative Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company		Mean Subunit Best Case		Mean Subunit Worst Case	
	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.550	0.550	0.433	0.433	0.679	0.679
50%	0.375	0.400	0.058	0.104	0.579	0.592
100%	0.213	0.250	0.013	0.017	0.433	0.433
200%	0.063	0.125	0.000	0.004	0.263	0.292
400%	0.000	0.025	0.000	0.004	0.108	0.125

Figure 4A. BVAR Model False Positive Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

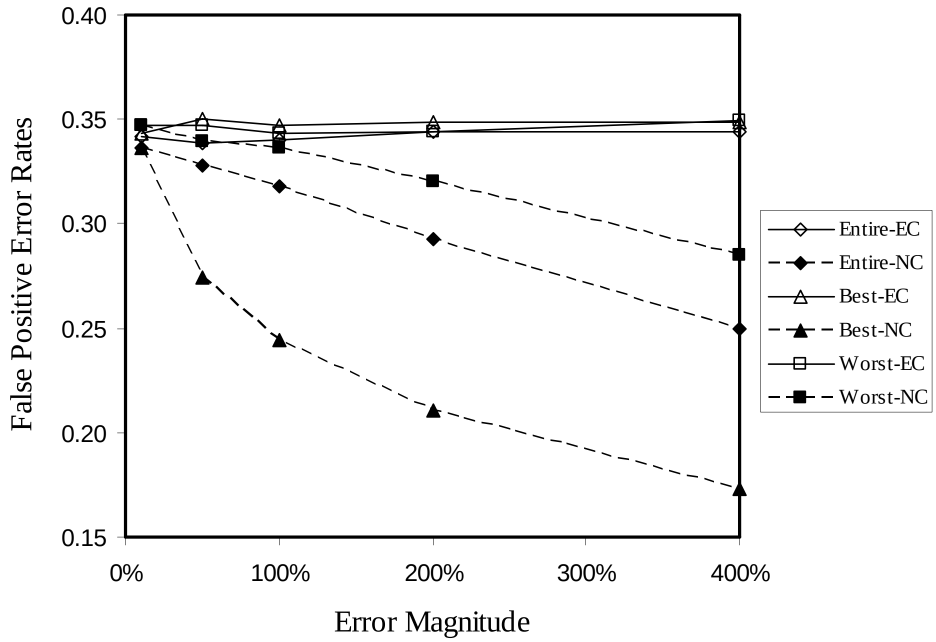
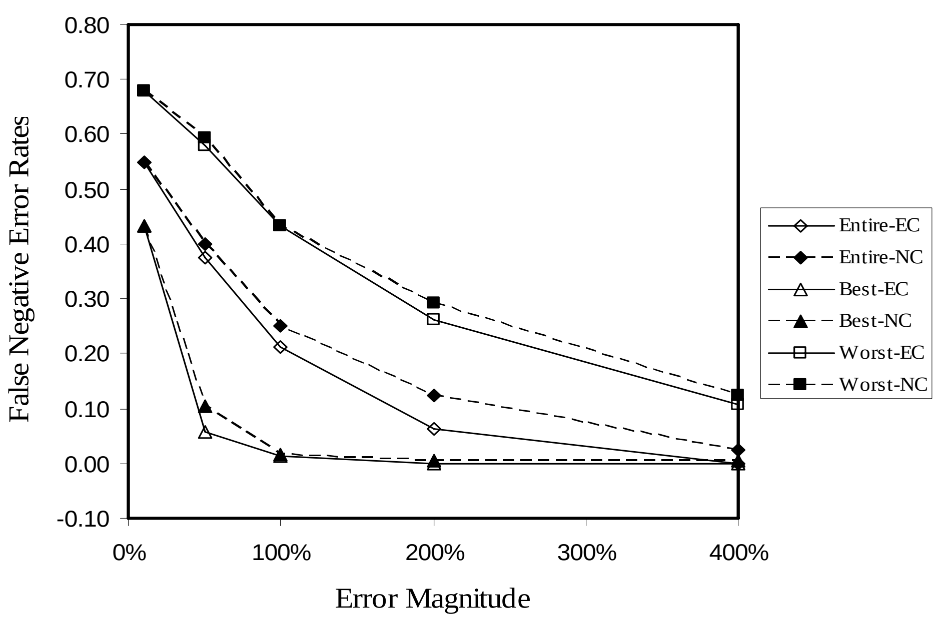


Figure 4B. BVAR Model False Negative Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)



Entire-EC: Error correction model using entire company data
Entire-NC: Non-correction model using entire company data
Best-EC: Error correction model using mean subunit best case scenario data
Best-NC: Non-correction model using mean subunit best case scenario data
Worst-EC: Error correction model using mean subunit worst case scenario data
Worst-NC: Non-Error correction model using mean subunit worst case scenario data

Table 15A: Subset VAR False Positive Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company		Mean Subunit Best Case		Mean Subunit Worst Case	
	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.039	0.039	0.032	0.032	0.032	0.032
50%	0.039	0.025	0.032	0.020	0.032	0.032
100%	0.033	0.014	0.032	0.020	0.032	0.032
200%	0.039	0.004	0.032	0.009	0.032	0.027
400%	0.039	0.002	0.032	0.006	0.032	0.021

Table 15B: Subset VAR False Negative Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company		Mean Subunit Best Case		Mean Subunit Worst Case	
	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.925	0.925	0.867	0.875	0.950	0.950
50%	0.825	0.825	0.433	0.513	0.913	0.913
100%	0.615	0.663	0.125	0.233	0.867	0.875
200%	0.212	0.350	0.000	0.117	0.667	0.742
400%	0.013	0.200	0.000	0.075	0.488	0.554

Figure 5A. Subset VAR Model False Positive Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

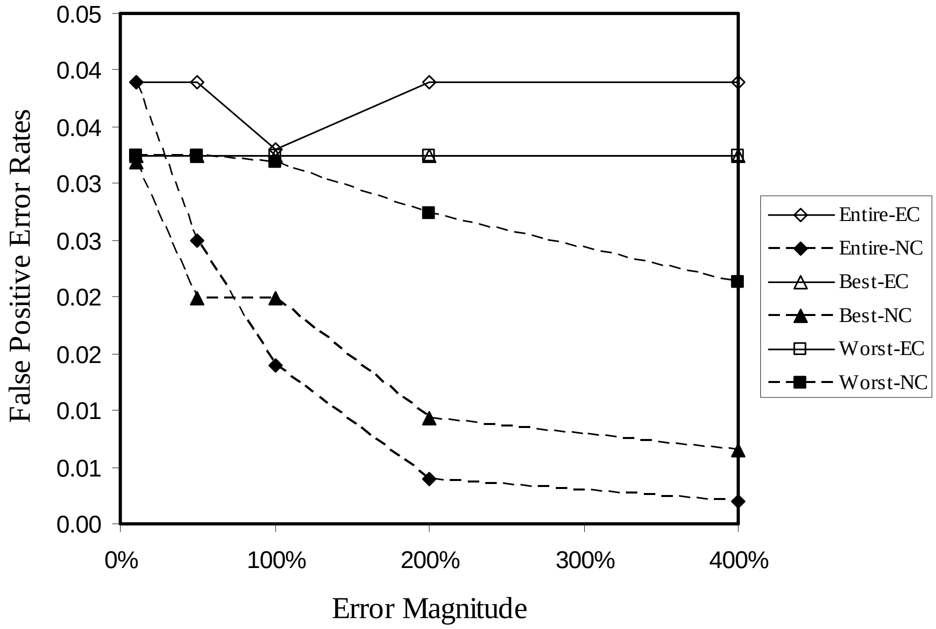
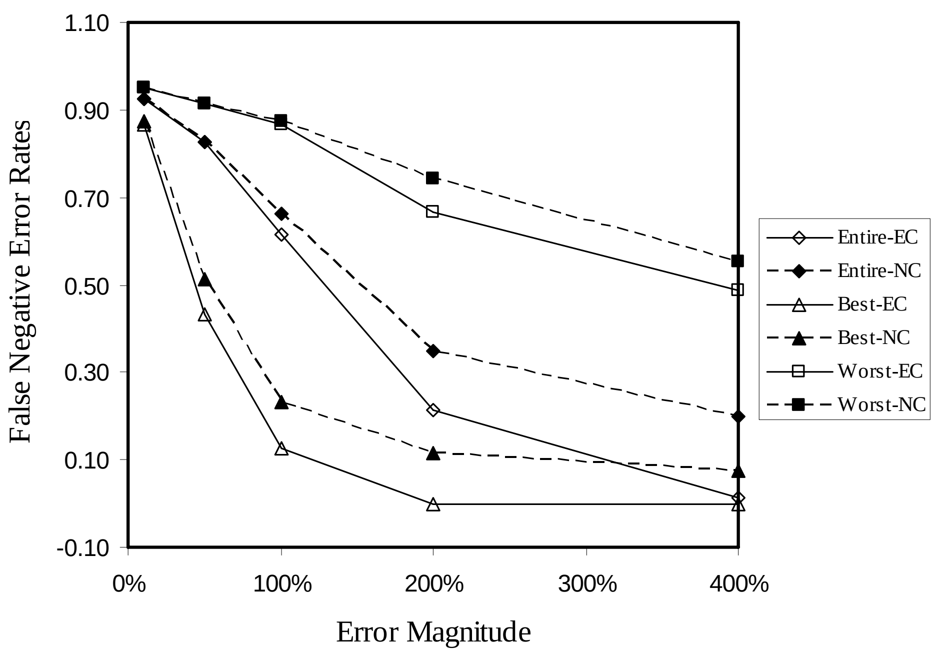


Figure 5B. Subset VAR Model False Negative Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)



Entire-EC: Error correction model using entire company data
Entire-NC: Non-correction model using entire company data
Best-EC: Error correction model using mean subunit best case scenario data
Best-NC: Non-correction model using mean subunit best case scenario data
Worst-EC: Error correction model using mean subunit worst case scenario data
Worst-NC: Non-Error correction model using mean subunit worst case scenario data

Table 16A: SEM False Positive Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company		Mean Subunit Best Case		Mean Subunit Worst Case	
	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.139	0.139	0.082	0.151	0.080	0.080
50%	0.134	0.120	0.080	0.091	0.081	0.081
100%	0.141	0.100	0.078	0.058	0.082	0.081
200%	0.140	0.063	0.075	0.021	0.082	0.073
400%	0.133	0.029	0.073	0.018	0.080	0.035

Table 16B: SEM False Negative Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company		Mean Subunit Best Case		Mean Subunit Worst Case	
	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.813	0.813	0.796	0.733	0.904	0.904
50%	0.775	0.775	0.254	0.221	0.863	0.875
100%	0.425	0.425	0.071	0.108	0.796	0.808
200%	0.188	0.250	0.013	0.054	0.613	0.638
400%	0.025	0.113	0.013	0.038	0.363	0.438

Figure 6A. SEM Model False Positive Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

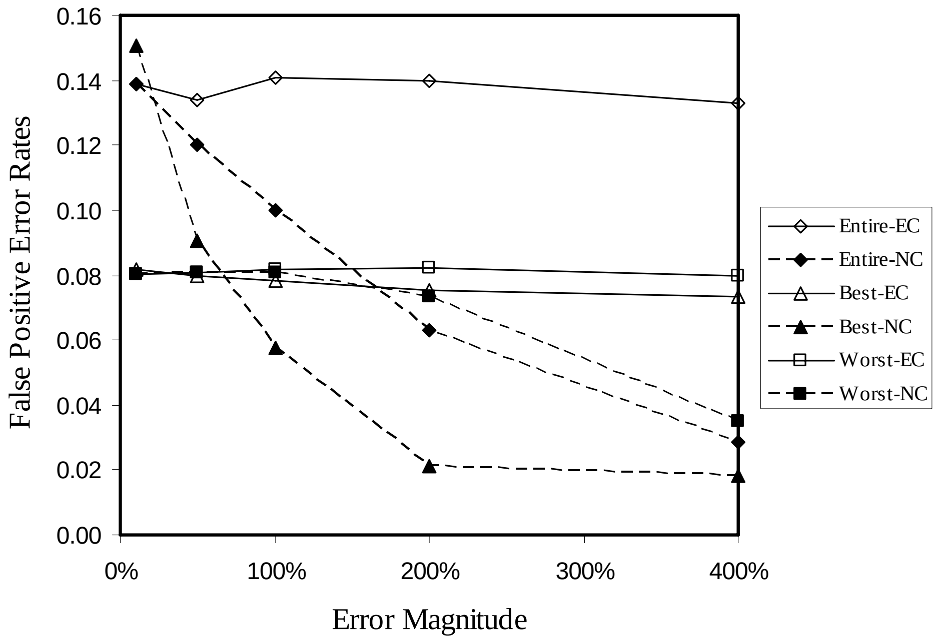
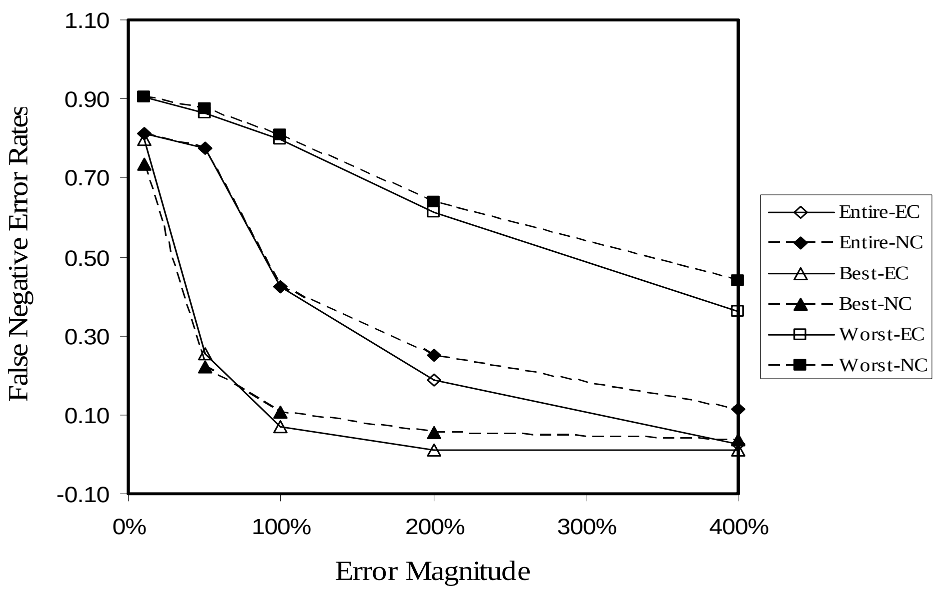


Figure 6B. SEM Model False Negative Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)



Entire-EC: Error correction model using entire company data
Entire-NC: Non-correction model using entire company data
Best-EC: Error correction model using mean subunit best case scenario data
Best-NC: Non-correction model using mean subunit best case scenario data
Worst-EC: Error correction model using mean subunit worst case scenario data
Worst-NC: Non-Error correction model using mean subunit worst case scenario data

Table 17A: LRM False Positive Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company		Mean Subunit Best Case		Mean Subunit Worst Case	
	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.090	0.090	0.058	0.053	0.058	0.058
50%	0.089	0.064	0.057	0.029	0.058	0.057
100%	0.089	0.046	0.058	0.025	0.058	0.053
200%	0.090	0.029	0.058	0.016	0.058	0.048
400%	0.090	0.020	0.058	0.010	0.057	0.034

Table 17B: LRM False Negative Error Rates Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

Error Magnitude	Entire Company		Mean Subunit Best Case		Mean Subunit Worst Case	
	Error-Correction	Non-Correction	Error-Correction	Non-Correction	Error-Correction	Non-Correction
10%	0.863	0.863	0.842	0.850	0.921	0.921
50%	0.813	0.813	0.333	0.400	0.904	0.904
100%	0.525	0.613	0.108	0.171	0.842	0.850
200%	0.188	0.288	0.000	0.071	0.642	0.667
400%	0.013	0.175	0.000	0.042	0.450	0.496

Figure 7A. LRM Model False Positive Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)

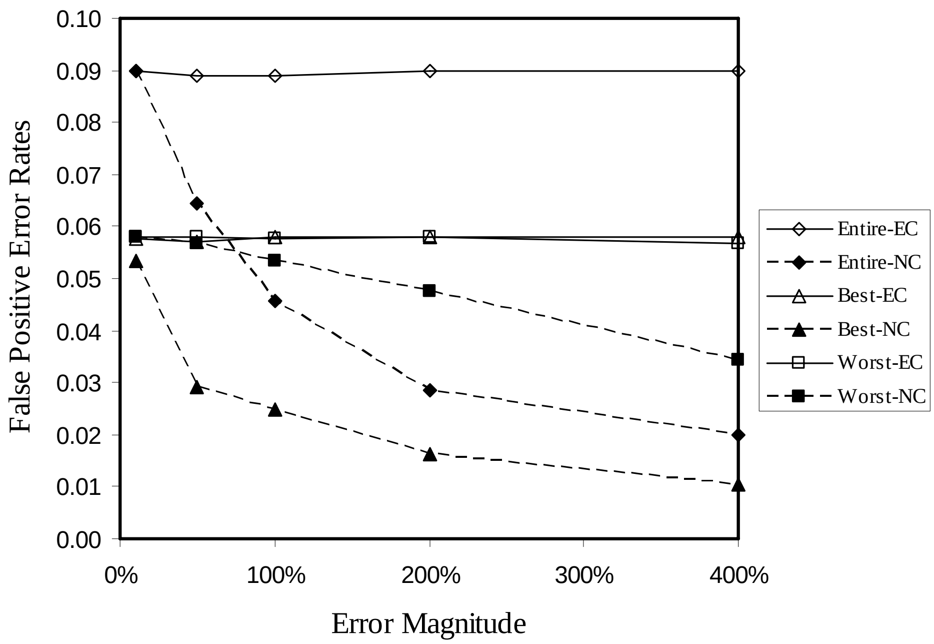
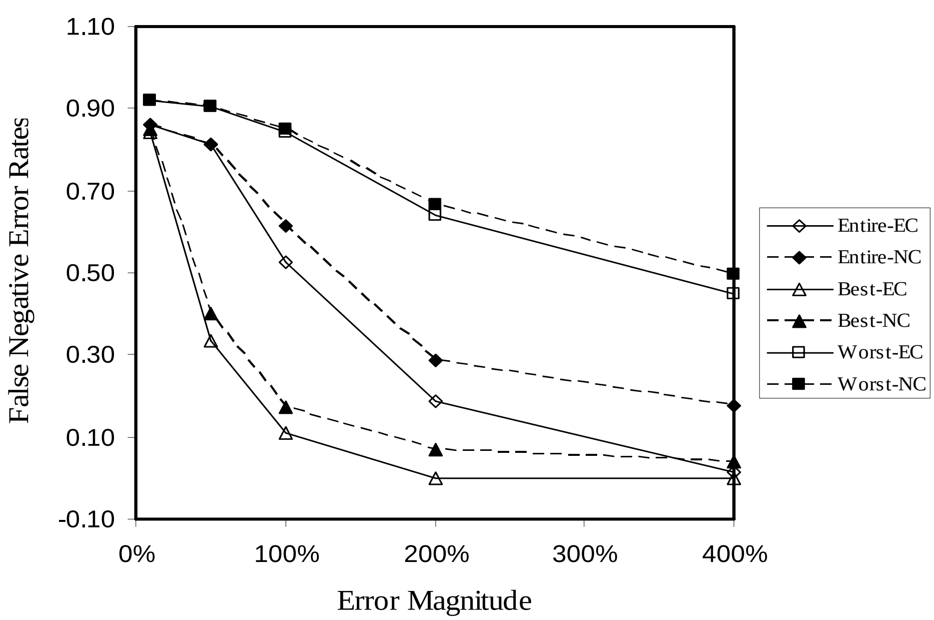


Figure 7B. LRM Model False Negative Rate Comparisons (Entire firm, Subunit Best Case, Subunit Worst Case)



Entire-EC: Error correction model using entire company data
Entire-NC: Non-correction model using entire company data
Best-EC: Error correction model using mean subunit best case scenario data
Best-NC: Non-correction model using mean subunit best case scenario data
Worst-EC: Error correction model using mean subunit worst case scenario data
Worst-NC: Non-Error correction model using mean subunit worst case scenario data

Table 8A: False Negative Error Rates of SEM, MTSM, and Linear regression

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%	90.00%	96.25%	95%
50%	78.75%	71.25%	68.75%
100%	33.75%	32.5%	33.75%
200%	12.50%	8.75%	17.5%
400%	0	0	2.5%

The false negative error rate indicates the percentage of errors that are not detected by the AP model. It is calculated as: (total number of undetected errors) / 8 (which is the number of seeded errors)*100%.

Table 8B: Detection Rates of SEM, MTSM, and Linear regression

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%E	10.00%	3.75%	5%
50%E	21.25%	28.75%	31.25%
100%E	66.25%	67.50%	66.25%
200%E	87.50%	91.25%	82.50%
400%E	100.00%	100.00%	97.50%

The detection rate indicates the percentage of errors that have been successfully detected. It is calculated as: 100% - False Negative Error Percentage.

Chart 4A: Anomaly Detection Comparison of SEM, MTSM and Linear Regression — False Negative Error Rate.

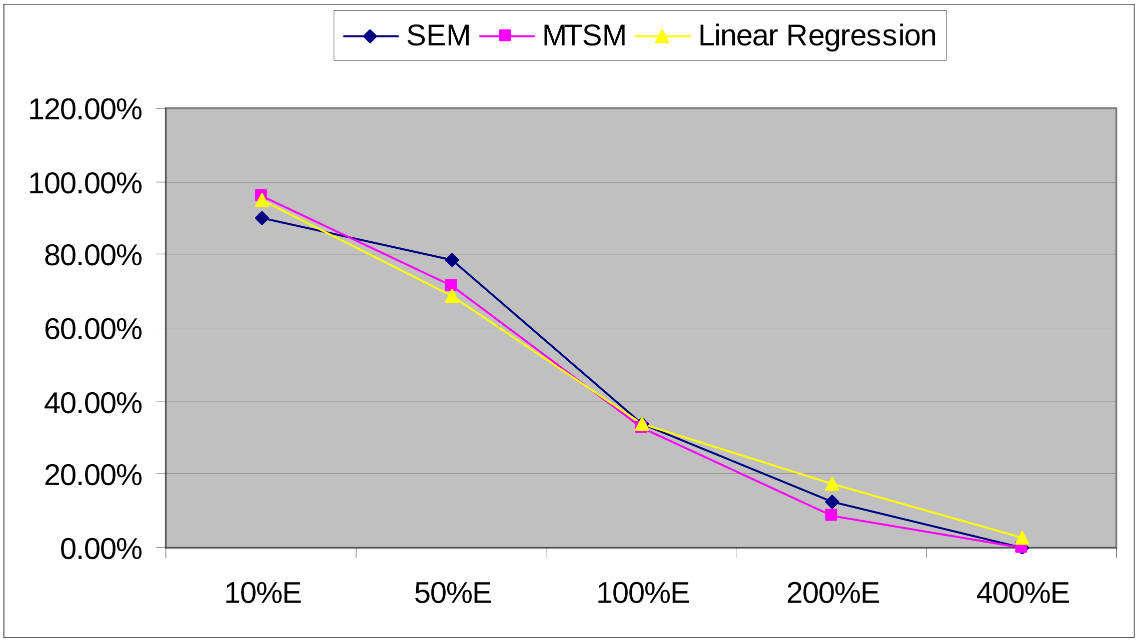


Chart 4B: Anomaly Detection Comparison of SEM, MTSM and Linear Regression — Detection Rate

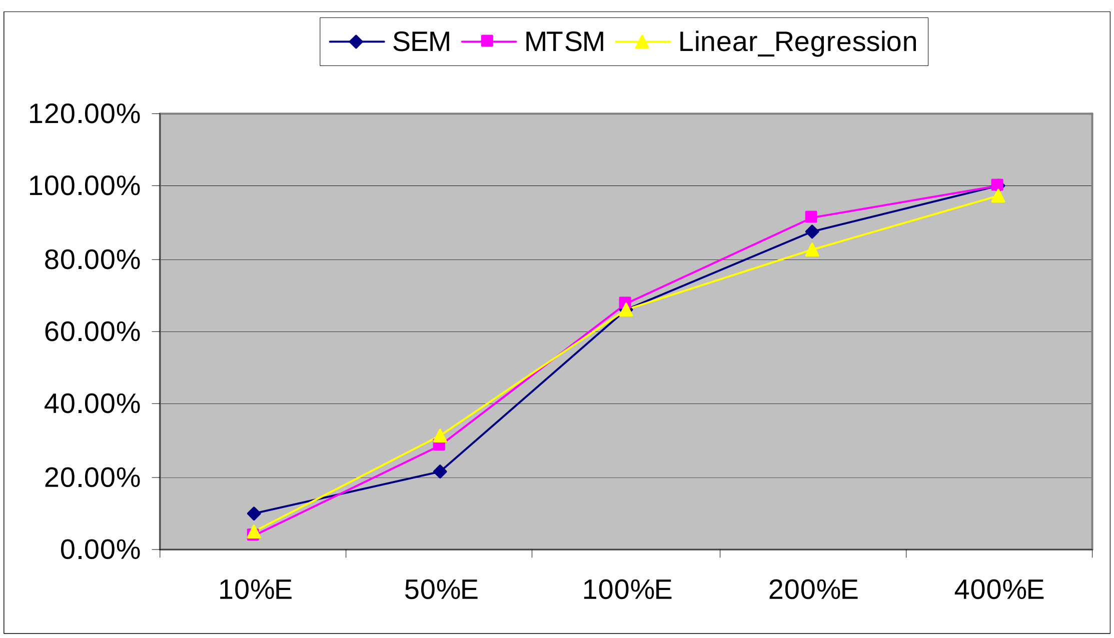


Table 9: False Positive Error Rates of SEM, MTSM, and Linear regression

Error Magnitude	SEM	MTSM	Linear Regression
10%	0	0	0
50%	0	0	0
100%	0	0	0
200%	0	0	0
400%	0	0	0

The false positive error rate indicates the percentage of non-errors that are reported by the AP model as errors. It is calculated as: (total number of reported non-errors) / (the number of observations in the hold-out set)*100%.

Table 10A: False Negative Error Rates of SEM, MTSM, and Linear regression for Tests Including Outliers and Error Correction.

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%	86.25%	86.25%	83.75%
50%	72.50%	75.00%	63.75%
100%	38.75%	37.50%	30.00%
200%	16.25%	12.50%	18.75%
400%	0.00%	0.00%	1.25%

The false negative error rate indicates the percentage of errors that are not detected by the AP model. It is calculated as: (total number of undetected errors) / 8 (which is the number of seeded errors)*100%.

Table 10B: Detection Rates of SEM, MTSM, and Linear regression for Tests Including Outliers and Error Correction.

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%	13.75%	13.75%	16.25%
50%	27.50%	25.00%	36.25%
100%	61.25%	62.50%	70.00%
200%	83.75%	87.50%	81.25%
400%	100.00%	100.00%	98.75%

The false negative error rate indicates the percentage of errors that are not detected by the AP model. It is calculated as: (total number of undetected errors) / 8 (which is the number of seeded errors)*100%.

Chart 5A: Anomaly Detection Comparison of SEM, MTSM and Linear Regression for Tests Including Outliers and Error Correction— False Negative Error Rate.

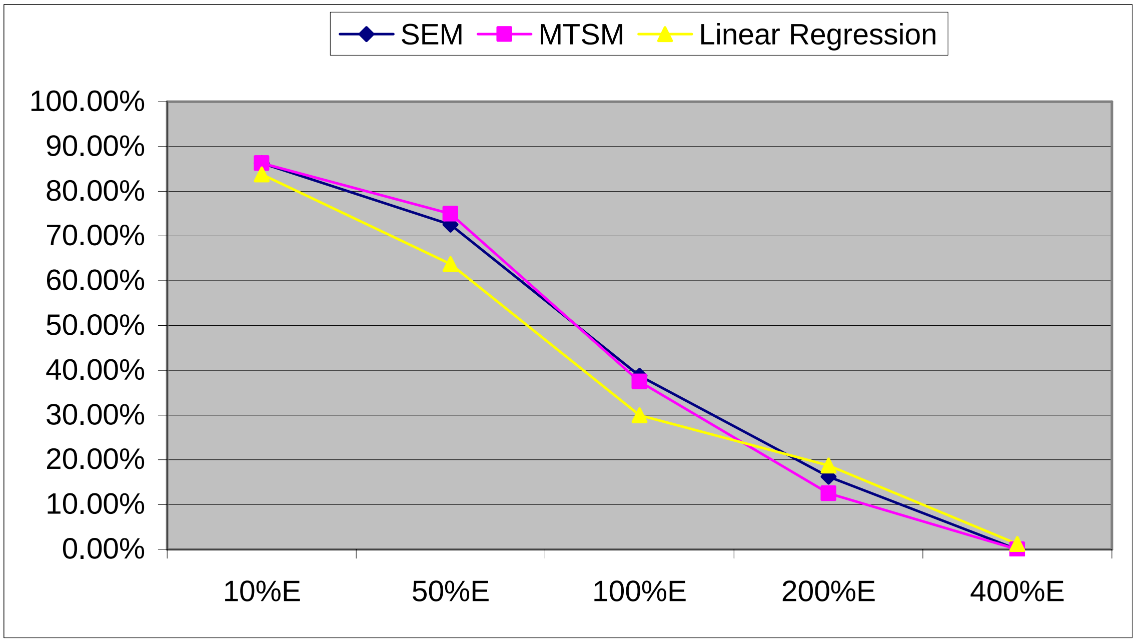


Chart 5A: Anomaly Detection Comparison of SEM, MTSM and Linear Regression for Tests Including Outliers and Error Correction— Detection Rate.

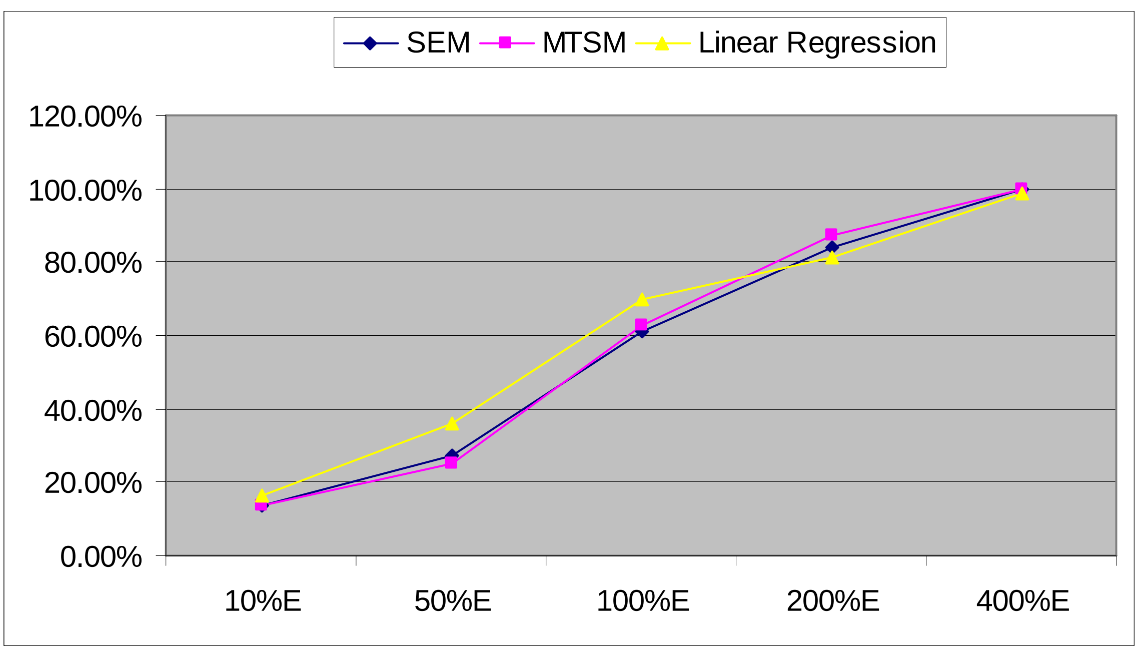


Table 11: False Positive Error Rates of SEM, MTSM, and Linear regression for Tests Including Outliers and Error Correction.

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%	10.00%	9.15%	9.78%
50%	10.00%	9.15%	9.78%
100%	10.00%	9.15%	10.00%
200%	10.00%	9.15%	9.78%
400%	10.00%	9.15%	9.78%

The false positive error rate indicates the percentage of non-errors that are reported by the AP model as errors. It is calculated as: (total number of reported non-errors) / 45 *100% for SEM and Linear Regression.; (total number of reported non-errors) / 47 *100% for MTSM.

Chart 6: Anomaly Detection Comparison of SEM, MTSM and Linear Regression for Tests Including Outliers— False Positive Error Rates.

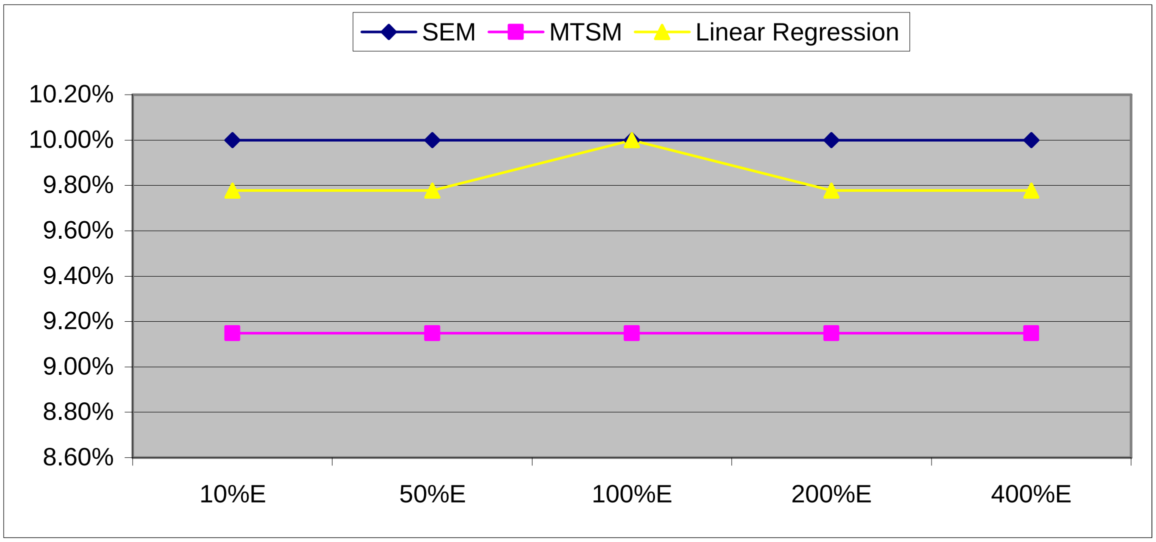


Table 12A: False Negative Error Rates of SEM, MTSM, and Linear regression for Tests Including Outliers and Without Error Correction.

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%	86.25%	86.25%	82.50%
50%	75.00%	76.25%	70.00%
100%	46.25%	42.50%	37.50%
200%	20.00%	17.50%	18.75%
400%	8.75%	12.50%	7.50%

The false negative error rate indicates the percentage of errors that are not detected by the AP model. It is calculated as: (total number of undetected errors) / 8 (which is the number of seeded errors)*100%.

Table 12B: Detection Rates of SEM, MTSM, and Linear regression for Tests Including Outliers and Without Error Correction.

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%	13.75%	13.75%	17.50%
50%	25.00%	23.75%	30.00%
100%	53.75%	57.50%	62.50%
200%	80.00%	82.50%	81.25%
400%	91.25%	87.50%	92.50%

The false negative error rate indicates the percentage of errors that are not detected by the AP model. It is calculated as: (total number of undetected errors) / 8 (which is the number of seeded errors)*100%.

Chart 7A: Anomaly Detection Comparison of SEM, MTSM and Linear Regression for Tests Including Outliers and Without Error Correction— False Negative Error Rate.

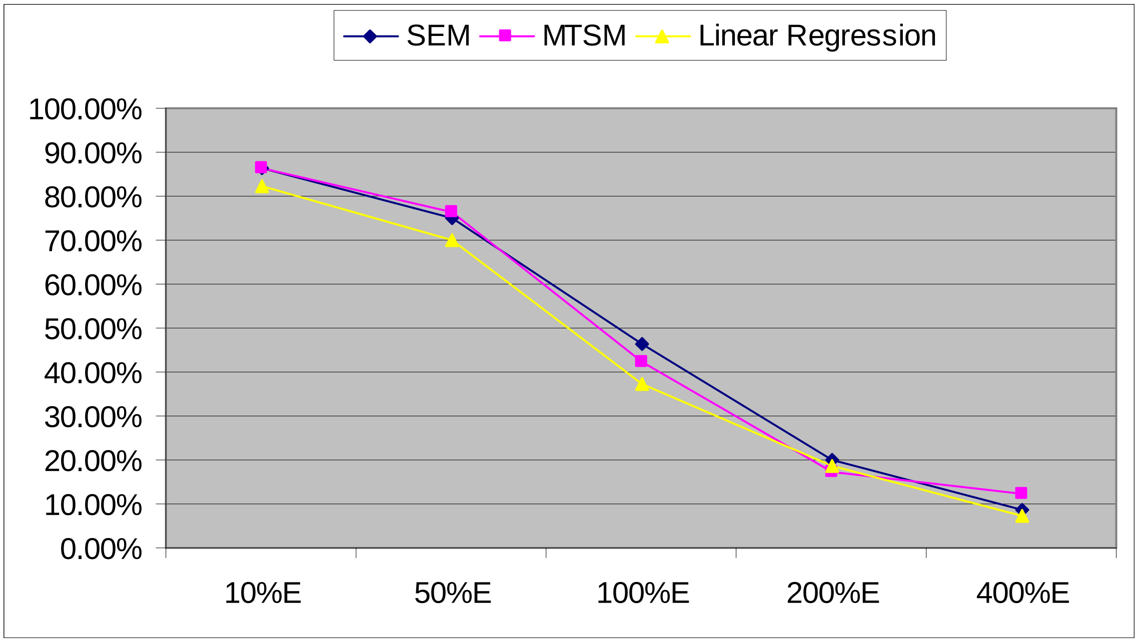


Chart 7B: Anomaly Detection Comparison of SEM, MTSM and Linear Regression for Tests Including Outliers and Without Error Correction— Detection Rate.

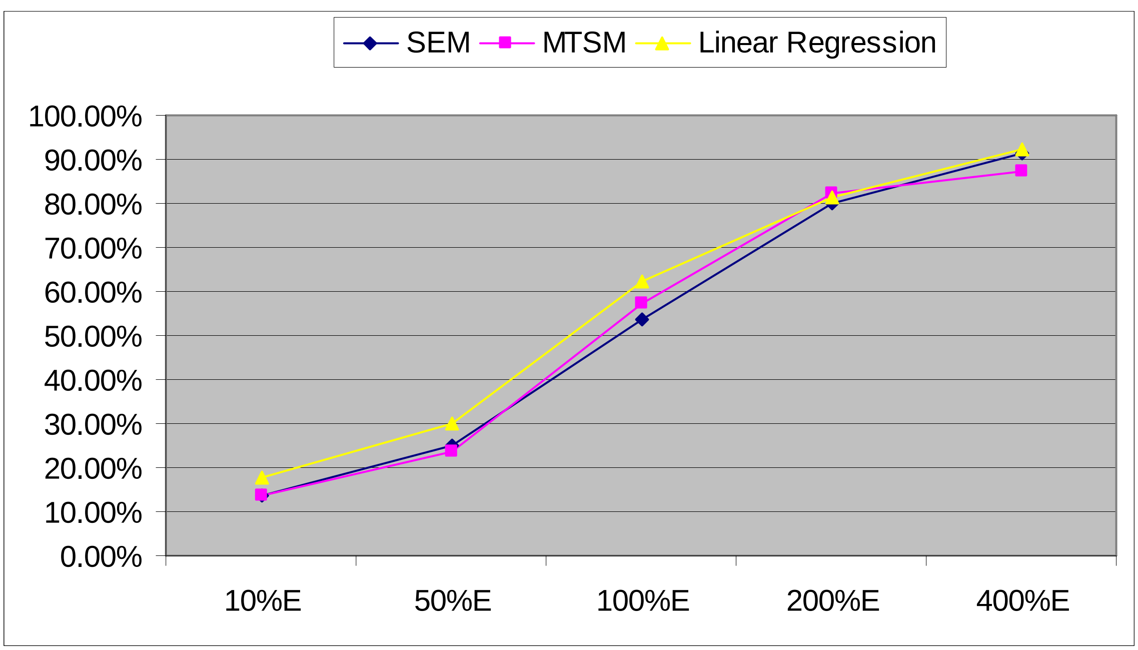


Table 13: False Positive Error Rates of SEM, MTSM, and Linear regression for Tests Including Outliers and Without Error Correction.

Error Magnitude	Simultaneous Equations	Multivariate Time Series	Linear Regression
10%	9.78%	9.15%	9.56%
50%	9.78%	8.09%	9.33%
100%	9.78%	6.81%	7.78%
200%	8.44%	5.74%	6.67%
400%	7.11%	4.04%	5.56%

The false positive error rate indicates the percentage of non-errors that are reported by the AP model as errors. It is calculated as: (total number of reported non-errors) / 45 *100% for SEM and Linear Regression.; (total number of reported non-errors) / 47 *100% for MTSM.

Chart 8: Anomaly Detection Comparison of SEM, MTSM and Linear Regression for Tests Including Outliers and Without Error Correction— False Positive Error Rates.