DATA REDUCTION TECHNIQUES AND HYPOTHESIS TESTING FOR ANALYSIS OF BENCHMARKING DATA

Jack A. Nickerson*

Thomas W. Sloan†

Rev. April 29, 2002

Abstract

This paper proposes a data reduction and hypothesis testing methodology that can be used to perform hypothesis testing with data commonly collected in benchmarking studies. A reduced-form performance vector and a reduced-form set of decision variables are constructed using the multivariate data reduction techniques of principal component analysis and exploratory factor analysis. Reductions in dependent and exogenous variables increase the available degrees of freedom, thereby facilitating the use of standard regression techniques. We demonstrate the methodology with data from a semiconductor production benchmarking study.

* John M. Olin School of Business, Washington University in St. Louis, Box 1133, One Brookings Drive, St. Louis MO 63130, USA. E-mail: *nickerson@wuolin.wustl.edu.*

† Department of Management, School of Business Administration, University of Miami, Coral Gables, FL 33124-9145, USA. E-mail: *tsloan@sba.miami.edu*, Tel: 305/284-1086, Fax: 305/284-3655. Please direct correspondence to this author.

1. INTRODUCTION

In less than two decades, benchmarking studies have become a mainstay for industry. Benchmarking studies attempt to identify relevant performance metrics and observe in great detail organizational and technological practices that lead to superior performance. In practice, however, identifying the factors that drive high performance, and in some instances identifying the performance metrics themselves, is problematic.

Systematically linking performance to underlying practices is one of the greatest challenges facing benchmarking practitioners and scholars alike. We conjecture that although benchmarking studies often produce a wealth of microanalytic data, identifying causal linkages is problematic for two reasons. First, practitioners often rely on inappropriate or ad hoc techniques for identifying the factors that underlie performance; these techniques are prone to biases and errors of many types. Even when relying on more systematic statistical methodologies, researchers frequently are unable to test hypotheses because of insufficient degrees of freedom (e.g., for hypothesis testing to take place the number of observations must exceed the sum of the number of statistical parameters being estimated). Second, identifying an appropriate set of performance metrics is often complicated by the fact that many metrics are inter-related in complex ways. How does one usefully analyze data collected in benchmarking efforts? How can hypotheses about which practices are efficiency enhancing and which ones are efficiency depleting be statistically examined? Or, more generally, how can we systematically identify the organizational practices critical to high performance?

This paper attempts to address these questions by proposing a methodology for systematically identifying linkages between performance metrics and organizational and technological decision variables that describe the various practices employed by firms when the number of observations is small. The approach is based on the multivariate data reduction techniques of principal component analysis and exploratory factor analysis. The methodology reduces the number of dependent (performance) variables by employing principal component

analysis to construct a reduced-form performance vector. Decision variables, whether technological or organizational, are grouped and reduced using exploratory factor analysis. Data reduction increases the available degrees of freedom thereby allowing the use of standard hypothesis testing techniques such as regression analysis.

After presenting the empirical methodology in more detail, we use it to analyze a benchmarking study in the semiconductor industry. The methodology is implemented using data gathered through the Competitive Semiconductor Manufacturing Study (CSMS) sponsored by the Alfred P. Sloan Foundation and undertaken by researchers at the University of California, Berkeley.

The paper proceeds as follows. Section 2 briefly describes the growth in benchmarking activities and reviews some of the extant data analysis approaches. Section 3 describes the proposed empirical methodology including a description of principal component analysis, factor analysis, and hypothesis testing. Section 4 applies the methodology to data provided by the CSMS, and Section 5 discusses advantages and limitations of the approach and plans for future work. Section 6 concludes.

2. BACKGROUND

Although firms have long engaged in many forms of competitive analysis, benchmarking is a relatively new phenomenon emerging only in the last 20 years. *Benchmarking* is the systematic study, documentation, and implementation of "best" organizational practices. Driving the growth of benchmarking is the view that best practices can be identified and, once identified, managers can increase productivity by implementing the best practice.

Benchmarking was introduced in the United States by Xerox. Faced with tremendous competitive challenges in the late 1970s and early 1980s from Japanese photocopier firms, Xerox began detailed studies of operations of their competitors as well as firms in related fields and developed a method for identifying best practices. By formulating and implementing plans based on identified best practices, Xerox was able to significantly improve its productivity,

performance, and competitive position. Once Xerox's success was recognized, other large corporations quickly followed suit. It was not until 1989, however, that the use of benchmarking greatly accelerated making it a mainstream business activity by firms of all sizes and industries.[1]

A contributing factor to the explosion of benchmarking activity was the publication of *The Machine that Changed the World* (Womack *et al*. 1990). This book reported on the International Motor Vehicle Program, a pioneering cooperative effort between academia, industry, and government, initiated by the Massachusetts Institute of Technology (M.I.T.). A multi-disciplinary and multi-institutional team of researchers studied over 35 automobile manufacturers, component manufacturers, professional organizations, and government agencies to identify variations in performance and the underlying factors that accounted for them. While the first phase of the study was completed between 1985 and 1990, the program continues today with an ever-increasing number of industry participants.

Recognizing the possible productivity gains that benchmarking efforts could provide to American industry, the Alfred P. Sloan Foundation initiated a program in 1990, the expenditures of which now total over $20 million, to fund studies of industries important to the U.S. economy. Industries currently under study include automobiles (M.I.T.), semiconductors (U.C. Berkeley), computers (Stanford), steel (Carnegie Mellon/University of Pittsburgh), financial services (Wharton), clothing and textiles (Harvard), and pharmaceuticals (M.I.T.). The program joins universities, which provide independent and objective research, with industry, which provides data, guidance, and realism. It is hoped that these studies will reveal a deeper understanding of those factors that lead to high manufacturing performance across a variety of industries and, ultimately, increase industrial productivity and fuel economic growth.

The benchmarking process employed by these studies is a variant of the standard process outlined in popular literature. The implicit model underlying this process is that performance is driven by a number of decision variables either implicitly or explicitly set by management. We

---

[1] Benchmarking literature has exploded in the last 15 years. A recent sample of the ABI/Inform database (a database of over 1,000 business-related journals) revealed that over 750 articles related to benchmarking have been written between 1974 and 1995. Over 650 of these were published *after* 1989.

assume the performance metrics are endogenous and the decision variables exogenous. The basic benchmarking process is summarized by the following four steps:[2]

1. Identify the underlying factors that drive performance.
2. Find "similar" firms, measure their performance, and observe their practices.
3. Analyze the data collected, compare performance to other firms, and identify and prioritize opportunities for improvement.
4. Develop and implement plans to drive improvement.

Steps 1, 2, and 3 are especially problematic for managers and researchers alike.[3] Correlating underlying practices with performance frequently has an indeterminate structure — the number of parameters to be estimated exceeds the degrees of freedom. The number of firms observed is generally small; much data is qualitative in nature; and the number of variables observed within each firm is large, making a statistical analysis nearly impossible.

Popular benchmarking literature says little about resolving this empirical issue. Instead of employing statistical analysis, practitioners reportedly rely on visual summaries of the data in the form of graphs and tables. For example, the Competitive Semiconductor Manufacturing Study (Leachman 1994, Leachman and Hodges 1996), which provides the data for the empirical analysis provided later in the paper, used visual summaries of the performance metrics to both describe data and draw inferences. The choice of which parameters to plot (which may heavily influence "observed" patterns) often relies on heuristics, intuitions, and guesses. Observing in a variety of plots the relative position of each firm under study presumably reveals which practices lead to high performance. Relying on approaches that do not provide statistical inference to

---

[2]See, for example, McNair and Leibfried (1992).

[3]We also note that identifying metrics that describe performance (i.e., not decision variables) is often difficult. Defining "good" performance is difficult because performance is typically multidimensional and involves tradeoffs. Is a firm that performs well along one dimension and poorly along a second dimension better-performing than a firm with the opposite performance characteristics? The methodology described in Section 3.1 provides some insight into the choice of performance metrics.

identify the correspondence between high performance and critical practices can lead to incorrect characterizations and, possibly, to decreases in productivity rather than to improvements.

Many researchers have attempted to go beyond graphical methods by exploring statistical associations between firm practices and performance. For instance, Powell (1995) used correlation analysis to shed light on the relationship between total quality management (TQM) practices and firm performance in terms of quality and competitiveness. He surveyed more than 30 manufacturing and service firms and found that adoption of TQM was positively related to several measures of financial performance. However, correlation analysis, like graphical approaches, lacks the ability to test specific hypotheses regarding the relationships between practices and performance.

Regression analysis is a common method for examining relationships between practices and performance and for testing hypotheses. For instance, Hendricks and Singhal (1996) employed regression analysis in their study of how TQM relates to financial performance for a broad range of firms. The authors found strong evidence that effective TQM programs (indicated by the receipt of quality awards) are strongly associated with various financial measures such as sales. While this study demonstrated the value of TQM programs in general, it did not attempt to identify links between specific practices and high performance. Furthermore, all of the performance measures were financial: sales, operating income, and operating margin. In many benchmarking studies, the performance measures of interest are not so clear-cut. Running simple regressions on individual performance metrics only tells part of the story, as each metric may only be a partial measure of some underlying performance variable. In many if not most cases, individual regressions will not reveal the relationship between practices and performance because the various performance metrics are related to each other in complex ways.

Another systematic approach employed to understand benchmarking data is data envelopment analysis (DEA), first proposed by Charnes *et al.* (1978). DEA assesses the relative efficiency of firms by comparing observed inputs and outputs to a theoretical production possibility frontier. The production possibility frontier is constructed by solving a set of linear

programs to find a set of coefficients that give the highest possible efficiency ratio of outputs to inputs.

DEA suffers from several drawbacks from the perspective of studying benchmarking data. First, DEA implicitly assumes that all the organizations studied confront identical production possibility frontiers and have the same goals and objectives. Thus, for firms with different production possibility frontiers, as in the semiconductor industry, DEA is neither appropriate nor meaningful. Second, performance is reduced to a single dimension, efficiency, which may not capture important learning and temporal dimensions of performance. Third, DEA by itself simply identifies relatively inefficient firms. No attempt is made to interpret performance with respect to managerial practices.

Jayanthi *et al.* (1996) went a step beyond DEA in their study of the relationship between a number of manufacturing practices and firm competitiveness in the food processing industry. They measured the competitiveness of 20 factories using DEA and a similar method known as operational competitiveness ratings analysis (OCRA). They also collected data on various manufacturing practices such as equipment and inventory policies. Based on regression analysis, they concluded that several practices were indeed related to their measure of operational competitiveness. While this is an important step toward linking firm practices and performance, they only compared firms along a single performance dimension.

Canonical correlation analysis (CCA) is another method used to explore associations between firm practices and performance. Using this technique, one partitions a group of variables into two sets, a predictor set and a response set. CCA creates two new sets of variables, each a linear combination of the original set, in such a way as to maximize the correlation between the new sets of variables. Sakakibara *et al.* (1996) collected data from more than 40 plants in the transportation components, electronics, and machinery industries. They used canonical correlation to study the effects of just-in-time practices (a set of six variables) on manufacturing performance (a set of four variables). Szulanski (1996) employed CCA to examine how firms internally transfer best-practice knowledge. The author collected data on more than 100 transfers in eight large firms. While it is an effective way to measure the strength

of the relationship between two sets of variables, canonical correlation does not provide a way to test specific, individual hypotheses regarding the original variables. In other words, it is impossible to "disentangle" the new sets of variables and draw conclusions about the original variables.

Structural equation modeling (SEM) and its relative, path analysis, are other statistical methods that have been used to examine cause-and-effect relationships among a set of variables. For example, Collier (1995) used SEM to explore the relationships between quality measures, such as process errors, and performance metrics, such as labor productivity, in a bank card remittance operation. The author succeeded in linking certain practices and performance measures, but no inter-firm comparisons were made. Ahire *et al.* (1996) examined data from 371 manufacturing firms. They used SEM to examine the relationships among a set of quality management constructs including management commitment, employee empowerment, and product quality. Fawcett and Closs (1993) collected data from more than 900 firms and used SEM to explore the relationship between several "causes"—such as the firm's globalization perception and the degree to which its manufacturing and logistics operations were integrated—and a number of "effects" related to competitiveness and financial performance. Unfortunately, SEM requires very large samples to be valid, which is a significant obstacle for most benchmarking studies.

The weaknesses of these approaches suggest that the analysis of benchmarking data could be improved by a methodology that (1) overcomes the obstacle of small sample size, (2) provides the ability to test specific hypotheses, and (3) enables researchers to find underlying regularities in the data while maintaining a separation between practice (cause) and performance (effect). None of the methods mentioned above satisfy these needs.

3. PROPOSED METHODOLOGY

The main statistical obstacle faced by benchmarking studies is that of insufficient degrees of freedom. The number of variables involved in relating practice to performance typically far

exceeds the number of observations. Also, identifying key performance metrics is problematic because performance is often multifaceted. The approach developed herein attempts to overcome these obstacles by employing data reduction techniques to reduce the number of endogenous performance metrics and the number of exogenous decision variables. Reducing both endogenous and exogenous variables increases the degrees of freedom available for regression analysis thereby allowing, in some instances, statistical hypothesis testing.

3.1. Data Reduction of Performance Variables

What is good performance? Simple financial measurements such as profitability, return on investment, and return on assets are all firm-level measures that could be used to identify good and bad performance. Unfortunately, these firm-level metrics are highly aggregated and are inappropriate for benchmarking efforts of less aggregated activities such as manufacturing facilities. Performance metrics will vary by the unit of analysis chosen and by industry, and thus a universal set of metrics can not be established for all benchmarking studies. Rather, performance metrics must be carefully selected for each study.

Since practitioners are capable of identifying appropriate performance metrics (our *endogenous* variables), our focus turns to techniques for summarizing performance metrics used in practice. Reducing the number of endogenous variables confronts several problems. First, performance changes over time and is usually recorded in a time series which may exhibit wide fluctuations. How are time series data appropriately summarized? Second, benchmarking participants may provide windows of observation of varying time spans. How are data of varying time spans best summarized? Third, firms may provide windows of observation that are non-contemporaneous. Firms are constantly changing their product mix, equipment sets, and production practices. If a firm's performance improves over time, more recent data would cast the firm's performance in a more favorable light. How should data be summarized to account for non-contemporaneous measurement?

We propose to resolve these issues in the following ways. First, we propose that the time series of each performance metric for each firm be summarized by simple summary statistics over a measurement window of fixed length. For this study we choose to summarize

performance metrics by the mean and average rate-of-change for each time series.[4] Mean values are easily calculated and, in essence, smooth variations in the data. Average rates-of-change are useful for identifying trends. Although rates-of-change are distorted by random fluctuations in the data, they are important indicators of learning taking place within the firm.[5] Indeed, in many high technology industries, the rate-of-change (rates) may be equally if not more important than the absolute magnitude of performance (mean).

Second, we resolve the problem of observation windows of varying length by choosing the maximum common window length and ignoring all but the most recent time series data. Identifying the maximum common window length truncates the data and thus reduces the total amount of information available for analysis. Information loss notwithstanding, employing uniform observations windows improves the consistency of inter-firm comparisons and greatly facilitates more systematic analysis.

Third, we propose no adjustment for non-contemporaneous measurement when endogenous variables are reduced. Instead, we construct a vector that indexes when observations are made and consider the vector as an exogenous variable when testing hypotheses. We discuss the approaches further in Section 3.3.

We propose to reduce the set of endogenous variables with principal component analysis. The purpose of principal component analysis is to transform a set of observed variables into a smaller, more manageable set that accounts for most of the variance of the original set of variables. Principal components are determined so that the first component accounts for the largest amount of total variation in the data, the second component accounts for the second largest amount of variation, and so on. Also, each of the principal components is orthogonal to

---

[4]We also conceive of instances where the standard deviations of the rates-of-change of performance metrics provide an important summary statistic. We do not employ the use of standard deviations in this study because of the high rates of change in the semiconductor industry. Standard deviations, however, could be readily incorporated into our methodology.

[5]As with any discrete time-series data, calculating a rate-of-change amplifies measurement noise and hence distorts the information. The signal-to-noise ratio can be improved by averaging the rate-of-change across several contiguous measurements. The number of observations to average must be selected judiciously: noise will not be attenuated if few observations are averaged and unobserved but meaningful fluctuations will be attenuated if too many observations are averaged.

(i.e., uncorrelated with) the others. We argue that principal component analysis is the most appropriate technique with which to reduce endogenous variables because it imposes no pre-specified structure on the data and operates to maximize the amount of variance described by a transformed, orthogonal set of parameters. The advantage of this latter condition is that the transformed variables that account for little of the variance can be dropped from the analysis, reducing the number of endogenous variables. We describe this process in more detail below.[6]

Each principal component is a linear combination of the observed variables. Suppose that we have p observations, and let X_j represent an observed variable, where $j = 1, 2, \ldots, p$. The ith principal component can be expressed as

$$PC_{(i)} = \sum_{j=1}^{p} w_{(i)j} X_j \ ,$$

subject to the constraints that

$$\sum_{j=1}^{p} w_{(i)j}^2 = 1 \quad \text{for } i = 1, 2, \ldots, p, \text{ and} \tag{1}$$

$$\sum_{j=1}^{p} w_{(k)j} w_{(i)j} = 0 \quad \text{for all } i > k \tag{2}$$

where the w's are known as *weights* or *loadings*. Eq. (1) ensures that we do not arbitrarily increase the variance of the PC's; that is, we choose the weights so that the sum of the variances of all of the principal components equals the total variance of the original set of variables. Eq. (2) ensures that each principal component is uncorrelated with all of the previously extracted principal components.

Input to the model is either the variance-covariance matrix or the correlation matrix of the observations. There are advantages to using each of these matrices; however, the correlation matrix is often used because it is independent of scale, whereas the variance-covariance matrix is not; we use the correlation matrix for this reason. The output of the model is the set of loadings

[6]Our discussion of principal component analysis is based on Dillon and Goldstein (1984).

(i.e., the *w*'s). Regardless of the choice of inputs, each loading is a function of the eigenvalues of the variance-covariance matrix of the observations.

A reduced-form set of endogenous variables is identified by eliminating those eigenvectors that account for little of the data's variation. When the goal is data reduction, it is common to retain the minimum number of eigenvectors that account for at least 80 percent of the total variation. In many instances, what initially consisted of many variables can be summarized by as few as two variables.

3.2. Data Reduction for Exogenous/Decision Variables

Firm performance is presumably driven by a number of decision variables either implicitly or explicitly set by management. Variables might include, for example, choice of market position, production technology, organizational structure, and organizational practices such as training, promotion policies, and incentive systems. In the semiconductor industry, for example, fabrication facilities (fabs) that produce dynamic random access memory (DRAMs) have a different market focus than fabs that produce application specific integrated circuits (ASICs). Cleanliness of a fab, old production technology versus new, hierarchical versus flat organization structures, and specialized versus generic training are all examples of measurable variables. Most variables are readily observable through qualitative if not quantitative measurements.

For purposes of analysis, decision variables are assumed to be exogenous. However, it is important to note that not all variables are perfectly exogenous. Technology decisions may be more durable than some organizational decisions. The former describe sunk investments in durable goods whereas the latter describe managerial decisions that might be alterable in the near term. Indeed, labeling organization variables as exogenous may be problematic since poor performance may lead managers to alter organizational decisions more quickly than technological decisions. Technology and organization variables are considered separately later in the paper because of this potential difference in the durability of decisions.

The data used in our analysis, however, suggest that both technology and organization variables are relatively stationary over the period during which performance is measured. Hence, exogenous variables tend to be represented by single observations rather than a time series. If, however, exogenous variables are represented by a time series, we recommend adopting the data summary techniques described in Section 3.1.

How should we reduce the set of exogenous variables? Whereas principal component analysis is recommended for dependent variables, we claim that exploratory factor analysis is a more appropriate data reduction technique for exogenous variables. While principal component analysis maximizes data variation explained by a combination of linear vectors, factor analysis identifies an underlying structure of latent variables.[7] Specifically, factor analysis identifies interrelationships among the variables in an effort to find a new set of variables, fewer in number than the original set, which express that which is common among the original variables. The primary advantage of employing factor analysis comes from the development of a latent variable structure. Products, technology, and production processes used in fabs and their organization are likely to be a result of underlying strategies. Identifying approaches and strategies is useful not only as a basis for explaining performance variations but also for linking product, technology, and production strategies to performance. Factor analysis provides a means for describing underlying firm strategies; principal component analysis offers no such potential relationship.

The common factor-analytic model is usually expressed as

$$\mathbf{X} = \mathbf{\Lambda f} + \mathbf{e} \tag{3}$$

where $\mathbf{X}$ is a p-dimensional vector of observable attributes or responses, $\mathbf{f}$ is a q-dimensional vector of unobservable variables called common factors, $\mathbf{\Lambda}$ is a $p \times q$ matrix of unknown constants called *factor loadings*, and $\mathbf{e}$ is a p-dimensional vector of unobservable error terms. The model assumes error terms are independent and identically distributed (*iid*) and are

[7]Other metric-independent multivariate approaches such as multidimensional scaling and cluster analysis also are available. See Dillon and Goldstein (1984) for explication of these approaches.

uncorrelated with the common factors. The model generally assumes that common factors have unit variances and that the factors themselves are uncorrelated.[8]

Since the approach adopted here is exploratory in nature, a solution, should it exist, is not unique. Any orthogonal rotation of the common factors in the relevant q-space results in a solution that satisfies Eq. (3). To select one solution, we embrace an orthogonal varimax rotation which seeks to rotate the common factors so that the variation of the squared factor loadings for a given factor is made large. Factor analysis generates vectors of factor loadings, one vector for each factor, and generates a number that typically is much less than the original number of variables. From the loadings we can construct a ranking in continuous latent space for each fab.

Common factors are interpreted by evaluating the magnitude of their loadings which give the ordinary correlation between an observable attribute and a factor. We follow a procedure suggested by Dillon and Goldstein (1984) for assigning meaning to common factors.[9]

Exploratory factor analysis suffers from several disadvantages. First, unlike principal component analysis, exploratory factor analysis offers no unique solution and hence does not generate a set of factors that is in some sense unique or orthogonal. The lack of a unique solution limits the procedure's generalizability to all situations. Second, any latent structure identified by the procedure may not be readily interpretable. Factor loadings may display magnitudes and signs that do not make sense to informed observers and, as a result, may not be easily interpretable in every case.

---

[8]Binary exogenous variables do pose problems for factor analysis. Binary variables have binomial distributions that depart from the assumption of normally distributed errors. In general, factor analysis will produce outputs when variables are binary although with a penalty in reduced robustness. An often described technique for improving robustness is to aggregate groups of similar binary variables and sum the responses so that an aggregate variable(s) better approximate a continuous variable.

[9]Dillon and Goldstein (1984, p.69) suggest a four step procedure. First, identify for each variable the factor for which the variable provides the largest absolute correlation. Second, examine the statistical significance of each loading noting that for sample sizes less than 100, the absolute value of the loading should be greater than 0.30. Third, examine the pattern factor loadings that contribute significantly to each common factor. Fourth, noting that variables with higher loadings have greater influence on a common factor, attempt to assign meaning to the factor based on step three.

These caveats notwithstanding, exploratory factor analysis may still prove to be the most appropriate tool for data reduction of at least some of the exogenous variables, depending on the researcher's goals. For example, perhaps a researcher's principal interest is in the organizational parameters, yet he or she desires to control for variations in technology. If so, then factor analysis can be applied to the technology parameters with the absence of a unique solution or difficulty in interpreting the factor having little impact on the final analysis of the organizational parameters.

3.3. Hypothesis Testing

Reductions in both endogenous and exogenous variables in many instances will provide a sufficient number of degrees of freedom to undertake hypothesis testing.[10] Regression analysis can be used to examine hypotheses about practices that lead to high (or low) performance.[11] Employing regression analysis requires, at a minimum, that the number of observations exceeds the number of variables in the model.[12] We proceed to describe one possible model for testing hypotheses assuming data reduction techniques have provided sufficient degrees of freedom.

Eq. (4) describes one possible hypothesis-testing model. In this model, a vector of dependent performance variables is expressed as a function of exogenous variables which we have divided into two classes: technology and organization. Specifically,

$$\mathbf{D} = \mathbf{T}\boldsymbol{\beta}_1 + \mathbf{H}\boldsymbol{\beta}_2 + e, \tag{4}$$

where **D** is the reduced-form vector of dependent performance variables, **T** is the reduced-form vector of technology variables, **H** is a reduced-form set of organization variables, and ***e*** is a vector of *iid* error terms. Ordinary least squares estimates the matrices of coefficients, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$,

---

[10]Of course, even after data reduction some studies will not yield sufficient degrees of freedom to allow hypothesis testing. Even when the proposed methodology fails to support hypothesis testing, both principal component and factor analysis are useful for revealing empirical regularities in the data. Structural revelations may be central to undertaking an improved and more focused benchmarking study.

[11]For a more detailed discussion of these and other techniques see, for example, Judge *et al*. (1988).

[12]The minimum number of degrees of freedom will depend on the statistical technique employed. Nevertheless, more is preferred to fewer degrees of freedom.

by minimizing the squared error term. The model seeks to explain variation in the reduced-form dependent variables by correlating them with the reduced-form exogenous variables. In this formulation, coefficients are evaluated against the null hypothesis using a student *t* distribution (*t*-statistics).

Regression analysis also provides an opportunity to consider the implications of non-contemporaneous measurement problems alluded to in Section 3.1. Evaluating the effects of non-contemporaneous measurement is accomplished by augmenting the vector of exogenous variables, either **T** or **H** or both, with a variable that indexes when observations are made. For example, a firm which offers the oldest observation window is indexed to 0. A firm whose observation window begins one quarter later is indexed to 1. A firm whose observation window begins two quarters after the first firm's window is indexed to 2, and so on. The estimated parameter representing non-contemporaneous measurement then can be used to evaluate whether or not performance is influenced by non-contemporaneous measurement.

4. APPLICATION OF THE METHODOLOGY

4.1. Competitive Semiconductor Manufacturing Study

Under sponsorship of the Alfred P. Sloan Foundation, the College of Engineering, the Walter A. Haas School of Business, and the Berkeley Roundtable on the International Economy at the University of California, Berkeley have undertaken a multi-year research program to study semiconductor manufacturing worldwide.[13] The main goal of the study is to measure manufacturing performance and to investigate the underlying determinants of performance.

The main phase of the project involves a 50-page mail-out questionnaire completed by each participant followed up by a two-day site visit by a team of researchers. The questionnaire quantitatively documents performance metrics and product, technology, and production process attributes such as clean room size and class, head counts, equipment counts, wafer starts, die

[13]See Leachman (1994) or Leachman and Hodges (1996) for a more complete description of the study.

yields, line yields, cycle times, and computer systems. During site visits researchers attempt to identify and understand those practices that account for performance variations by talking with a cross section of fab personnel.

4.2. Performance Metrics

The Competitive Semiconductor Manufacturing Study (CSMS) identifies seven key performance metrics described briefly below. Variable names used in our analysis appear in parentheses.

- *Cycle time per layer* (CTPL) is defined for each process flow and measures the average duration, expressed in fractional working days, consumed by production lots of wafers from time of release into the fab until time of exit from the fab, divided by the number of circuitry layers in the process flow.

- *Direct labor productivity* (DLP) measures the average number of wafer layers completed per working day divided by the total number of operators employed by the fab.

- *Engineering labor productivity* (ENG) measures the average number of wafer layers completed per working day divided by the total number of engineers employed by the fab.

- *Total labor productivity* (TLP) measures the average number of wafer layers completed per working day divided by the total number of employees.

- *Line yield* (LYD) reports the average fraction of wafers started that emerge from the fab process flow as completed wafers.

- *Stepper throughput* (STTP) reports the average number of wafer operations performed per stepper (a type of photolithography machine) per calendar day. This is an indicator of overall fab throughput as the photolithography area typically has the highest concentration of capital expense and is most commonly the long-run bottleneck.

- *Defect density* (YDD) is the number of fatal defects per square centimeter of wafer surface area. A model, in this case the Murphy defect density model, is used to convert actual die yield into an equivalent defect density.

This paper contains benchmarking data from fabs producing a variety of semiconductor products including DRAMs, ASICs, microprocessors, and logic. For this paper, we obtained a complete set of observations for 12 fabs. Prior to employing principal component analysis, data is normalized and averaged to report a single mean and a single average rate-of-change for each

metric for each fab. When fabs run multiple processes, we calculate the average metric across all processes weighted by wafer starts per process. Means for each metric are calculated across the most recent 12-month period for which data exists. Average quarterly rates-of-change (rates) are calculated by averaging rates of improvement over the most recent four quarters. For some fabs, defect density is reported for a selection of die types. A single average defect density and rate-of-change of defect density is reported by averaging across all reported die types. The above process yields a total of 14 metrics, seven means and seven average rates-of-change. Note that rate of change for variables is designated by the prefix R.

Mean performance metrics for each fab along with summary statistics are reported in Table 1A. Table 2A reports average rates-of-change for performance metrics for each fab and summary statistics. Tables 1B and 2B provide correlation matrices for performance metrics and average rates of change, respectively.[14]

4.3. Product, Technology, and Production Variables

The CSMS reports several product, technology, and production variables. We adopt these variables as our set of exogenous variables. The 11 exogenous variables are described below. The variable names in parentheses correspond to the names that appear in the data tables at the end of the paper.

- *Wafer Starts* (STARTS) reports the average number of wafers started in the fab per week.
- *Wafer size* (W_SIZE) reports the diameter in inches (1 inch ≈ 0.0254 m) of wafers processed in the fab.
- *Number of process flows* (FLOWS) counts the number of different sequences of processing steps, as identified by the manufacturer that implemented in the fab.

[14] Note that several of the variables in Tables 1B and 2B are highly correlated. For instance, TLP with DLP and STTP with TLP in Table 1B and R_DLP with R_ENG, R_TLP, and R_STTP and R_ENG with R_TLP and R_STTP in Table 2B. The high correlation is expected because all of these metrics have in their numerator the average number of wafer layers completed per day. Unlike regression analysis, highly correlated variables are not problematic for the principal component procedure and, instead, are desirable because high correlation leads to a smaller number of transformed variables needed to describe the data.

- *Product type* (P_TYPE) identifies broad categories of products produced at a fab and is coded as 1 for memory, 0 for logic, and 0.5 for both.

- *Number of active die types* (D_TYPE) counts the number of different die types produced by a fab.

- *Technology* (TECH) refers to the minimum feature size of die produced by the most advanced process flow run in the fab. This is measured in microns (1 micron = 10^{-6}m).

- *Process Age* (P_AGE) refers to the age, in months, of the process technology listed above.

- *Die Size* (D_SIZE) is the area of a representative die type, measured in cm^2 ($1cm^2 = 10^{-4}m^2$).

- *Facility size* (F_SIZE) is the physical size of the fab's clean room. Small fabs with less than 20,000 ft^2 are coded as -1, medium size fabs with between 20,000 ft^2 and 60,000 ft^2 are coded as 0 and large fabs with more than 60,000 ft^2 are coded as 1 ($1\ ft^2 \approx 0.093\ m^2$).

- *Facility class* (CLASS) identifies the clean room cleanliness class. A class *x* facility has no more than 10^x particles of size 0.5 microns or larger per cubic foot of clean room space ($1\ ft^3 \approx 0.028\ m^3$).

- *Facility age* (F_AGE) identifies the vintage of the fab with pre-1985 fabs coded as -1, fabs constructed between 1985 and 1990 coded as 0, and fabs constructed after 1990 coded as 1.

Parameter values for the 11 exogenous technology variables along with summary statistics are reported in Table 3A. Table 3B reports the correlation matrix.[15]

4.4. Principal Component Analysis

We performed principal component analysis separately on the metric means and rates.[16] We first summarize the principal components of the means (shown in Table 4A), then summarize principal components of the rates (shown in Table 4B).

---

[15] Note that Table 3B shows that TECH and W_SIZE are highly correlated, which suggests that small circuit feature size corresponds to large wafer size. While the relationship is expected, it indicates that the variance in once variable is not perfectly accounted for by the other variable. Thus, it is appropriate for variables to remain in the factor analysis.

[16] Separate principal component analyses allow for closer inspection of performance rates-of-change as distinct from means. Both data sets were merged and collectively analyzed via principal component analysis with no change in the total number of principal components (five) and little variation in vector directions and magnitudes. For economy, the joint analysis is not reported.

Principal component analysis of the performance metric means shows that 83 percent of variation is described by two eigenvectors which we label M_PRIN1 and M_PRIN2. The third largest eigenvalue and its corresponding eigenvector describes less than nine percent additional variation, thus we conclude that the seven metrics describing mean performance levels over a one-year time period can be reduced to two dimensions. Component loadings and eigenvalues for the seven metrics are given in Table 4A.

We can describe the two eigenvectors by looking at the magnitude and sign of the loadings given in Table 4A. The loadings for eigenvector M_PRIN1 except for the one associated with defect density are similar in magnitude. The loading suggests that fabs that rank highly along this dimension display low cycle time (note the negative coefficient), high labor productivity of all types, high line yields, and high stepper throughput. Low cycle time allows fabs to respond quickly to customers and high labor productivity of all types, high line yields, and high stepper throughput corresponds to fabs that are economically efficient. We label component M_PRIN1 as a measure of *efficient responsiveness*.

We label eigenvector M_PRIN2 as a measure of *mass production*. This dimension is dominated by a negative correlation with defect density, i.e., low defect density yields a high score. Both cycle time, which has a positive coefficient, and engineering labor productivity, which has a negative coefficient, also strongly correlate with this dimension. Thus, eigenvector M_PRIN2 will yield a high score for fabs that have low defect densities, long cycle times, and low engineering productivity (i.e., more engineering effort). Fabs corresponding to these parameters are typically engaged in single-product mass production. For example, competitive intensity in the memory market leads DRAM fabs to focus on lowering defect density, which requires high levels of engineering effort even to produce small reduction in defect density, and maximizing capacity utilization, which requires buffer inventories for each bottleneck piece of equipment and leads to long cycle time.

Principal component analysis of the rate metrics shows that 92 percent of variation is described by the first three eigenvectors with the first eigenvector accounting for the lion's share (58 percent) and the second and third eigenvectors accounting for 18 percent and 16 percent of

the variation, respectively. The fourth largest eigenvalue (and its corresponding eigenvector) describes less than six percent additional variation, thus we conclude that the data is appropriately reduced to three dimensions which we label R_PRIN1, R_PRIN2, and R_PRIN3.

We label eigenvector R_PRIN1 as a measure of *throughput improvement* or capacity-learning-per-day. The weights for all three labor productivity rates are large and positive as is that for the rate-of-change of stepper throughput, which means wafer layers processed per day is increasing and that labor productivity is improving. The weight for rate-of-change for cycle time is large and negative, which means fabs receiving a high scoring are reducing cycle time.

We label eigenvector R_PRIN2 as a negative measure of *defect density improvement* or "just-in-time" learning. Positive and large coefficients for defect density and cycle time per layer suggest that increases in defect density go hand-in-hand with increases in cycle time. Or, viewed in the opposite way, decreases in defect density come with decreases in cycle time per layer at the cost of a small decrease in stepper throughput as is suggested by its small and negative coefficient. Note that high-performing fabs (high reductions in defect density and cycle time) receive low scores along this dimension while poorly performing fabs receive high scores.

We label eigenvector R_PRIN3 as the *line yield improvement* or line yield learning. Large improvements in line yield and to a lesser extent increases in cycle time and decreases in defect density contribute to high scores on this component.

4.5. Factor Analysis

Using factor analysis, we are able to reduce the 11 exogenous variables to four common factors. Table 5A reports the 11 eigenvalues for the technology metrics. The first four eigenvalues combine to account for 79 percent of the variation. With the fifth eigenvalue accounting for less 10 percent of the variation, the factor analysis is chosen to be based on four factors. Table 5B reports factor loadings and the variance explained by each factor. After rotation, the four common factors combine to describe approximately 79 percent of the total variation with the first factor describing approximately 25 percent, the second factor describing 23 percent, the third factor describing 17 percent, and the fourth factor describing 15 percent.

Each of the four factors can be interpreted by looking at the magnitude and sign of the loadings that correspond to each observable variable as described in Section 3.2. Referring to the loadings of the rotated factor pattern in Table 5B, Factor 1 is dominated by three variables: wafer size, technology (minimum feature size), and die size. A negative sign on the technology variable suggests that larger line widths decrease the factor score. Fabs that process large wafers, small circuit geometries, and large dice will have high values for Factor 1. In practice, as the semiconductor industry has evolved, new generations of process technology are typified by larger wafers, smaller line widths, and larger dice. Thus, we label Factor 1 as a measure of *process technology generation* with new process technology generations receiving high Factor 1 scores and old generations receiving low scores.

Factor 2 is strongly influenced by wafer starts and facility size and, to a lesser degree, by the number of process flows and the type of product. Specifically, large fabs that produce high volumes, have many different process flows, and emphasize memory products will receive high Factor 2 scores. Conversely, small fabs that produce low volumes, have few process flows, and emphasize logic (including ASICs) will receive a low Factor 2 score. We label Factor 2 as a measure of *process scale and scope*.

Factor 3 is dominated by process age, facility age, and, to a lesser degree, by product type. The older the process and facility, the higher the Factor 3 score. Also, a negative sign on the product type loading suggests that logic producers will have high scores for this factor. Old logic fabs will score highly in Factor 3 which we label as *process and facility age*.

Factor 4 is dominated by one factor: number of active die types. Thus, we label Factor 4 as *product scope*. Firms with many die types, such as ASIC manufacturers, will receive high Factor 4 scores.

4.6. What Drives Performance?

In order to illustrate the proposed methodology, we investigate the relationship between the reduced-form exogenous factors and the reduced-form performance metrics. Specifically, we evaluate the degree to which the reduced-form technology metrics of product, technology, and

production process influence a fabrication facility's reduced-form performance metrics by performing a series of regressions. In each regression, a reduced-form performance metric is treated as the dependent variable, and the reduced-form exogenous factors are treated as the independent variables. Organization variables are not included in our analysis. Also, we investigate the effects of non-contemporaneous measurement by constructing a vector that indexes when the observations were made, and treating this as an independent variable.[17] In each regression, the null hypothesis is that the reduced-form performance metric is not associated with the reduced-form exogenous factors (including the time index).

Evaluation of these hypotheses provides insight into the degree to which product, technology, and production process decisions influence fab performance. Or, put differently, we evaluate the degree to which these factors do not explain performance. Two sets of regressions are undertaken. Columns (1) and (2) in Table 6 report regression results for the two principal components describing reduced-form mean performance metrics. Columns (3), (4), and (5) in Table 6 report regression results for the three principal components describing reduced-form rate-of-change of performance metrics.

4.6.1. Analysis of Reduced-Form Mean Performance Metrics

Column (1) reports coefficient estimates for M_PRIN1 (efficient responsiveness). Only one variable, Factor 2 (process scale and scope), is statistically significant. This finding supports the proposition that firms that score high on process scale and scope display high degrees of efficient responsiveness. Note that this finding is generally consistent with the view that fabs making a variety of chips using a variety of processes compete on turn-around time, which is consistent with efficient responsiveness, instead of on low cost achieved through mass

[17]The most recent quarter of data collected from the 12 fabrication facilities falls within a two-year window between the beginning of 1992 and the end of 1993. The data selected for analysis is the last complete year of observations; the maximum temporal measurement difference is seven quarters. Since differences are measured in quarters after the first quarter of 1992, the measurement interval vector contains elements that vary between zero and seven in whole number increments.

production. The model produces an adjusted R^2 of 0.47 but the F statistic is insignificant, which suggests the independent variables may not have much explanatory power.

Regression analysis of the M_PRIN2 (mass production) shown in column (2) suggests that the independent variables provide a high degree of explanatory power. The model has an adjusted R^2 of 0.71 and an F value that is statistically significant. Two parameters, Factor 1 (process technology generation) and Factor 3 (process and facility age), have coefficients that are statistically sufficient. We can interpret the coefficients as suggesting that new generations of process technology and young processes and facilities are used for mass production. Indeed, this result supports the commonly held view that high-volume chips such as DRAMS are technology drivers, which drive both the introduction of new technology and the construction of new facilities. In both regressions, we note that non-contemporaneous measurement has no significant effect.

These two regressions suggest that the mean performance metrics are related to technology metrics — that is, the choice of technology predicts mean performance levels. Importantly, if the choice of technology reflects a firm's strategic position (e.g., a DRAM producer focused on mass production of a single product compared to an ASIC producer focused on quick turn-around of a wide variety of chips produced with a variety of processes) then benchmarking studies must control for the fact that firms may pursue different strategies by adopting different technologies.

4.6.2. Analysis of Reduced-Form Rate-of-Change Performance Metrics

The regression analysis for R_PRIN1 (throughput improvement) is shown in column (3). The analysis shows that none of the independent variables are statistically significant. Moreover, neither the adjusted R^2 nor the F statistic suggest a relationship between the reduced-form technology factors and throughput improvement. This result suggests that factors other than technology, perhaps organizational factors, are the source of throughput improvements.

Similarly, regression analysis of R_PRIN2 (column (4)) provides little support for a relationship between technology and defect density improvement. Only Factor 3, process and

facility age, is significant, but at the 90-percent confidence interval. The relationship suggests that new processes and facilities correspond to high rates of defect density improvement. The low adjusted R^2 and insignificant F statistics suggest that other factors are responsible for improvements in defect density.

Unlike the prior two models, the regression model for R_PRIN3 (line yield improvement), shown in column 5, does indicate a relationship between technology and performance improvement. Line yields improve with (1) new process technology (although only weakly), (2) small fabs that employ few process flows (process scale and scope), and (3) greater product variety (product scope). The model yields an adjusted R^2 of 0.65 and an F value that is statistically significant. The result can be interpreted with respect to the type of fab. Custom ASIC fabs (because they produce many products with few processes) with relatively new process technology experience the greatest line yield improvements.[18] Note that from a strategic standpoint, improving line yield is more important to ASIC fabs than other fabs because wafers broken during processing impose not only high opportunity costs (because of customer needs for quick turn around) but also could potentially damage their reputation for quick turn-around.

In summary, the three regression models predicting rates of improvements provide an insight into performance not revealed by the regressions involving the reduced-form mean performance metrics. Except for Factor 3 in the second equation, none of the independent variables influence the rate-of-change for R_PRIN1 and R_PRIN2. Variations in the rate-of-change for these two components appear to be a result of other factors not included in the model. Variations in the rate-of-change for the third component, R_PRIN3, are explained to a high degree by Factors 1, 2, and 4.

[18] Interestingly, this finding is consistent with the observation that some of the older ASIC fabs studied introduced a new production technology for handling wafers, which greatly reduced wafer breakage.

5. DISCUSSION

The Competitive Semiconductor Manufacturing study provides an interesting opportunity for evaluating the proposed methodology. Without employing data reduction techniques, the study must grapple with twelve complete observations, seven performance metrics, and at least eleven exogenous variables describing variations in products, technologies, and production processes.[19] The unreduced data offer no degrees of freedom for testing hypotheses relating practices to performance. The methodology developed in this paper and applied to the CSMS data shows promise for resolving the data analysis challenges of benchmarking studies in general.

Application of principal component analysis reduced seven performance metrics (fourteen after time series data is summarized by means and rates-of-change) to five reduced-form variables. Factor analysis reduced technology variables from eleven to four. Whereas regression analysis initially was impossible, data reduction allowed our six-variable model to be analyzed with six degrees of freedom (twelve observations less six degrees of freedom for the model).

Regression analysis indicates that while reduced-form technology variables greatly influence the mean level of performance, they have a limited impact in explaining variations in the rate-of-change of performance variables. Clearly, other factors such as organizational practices are likely to be driving performance improvements. Indeed, analysis of the reduced-form data provides a baseline model for evaluating alternative hypotheses since it provides a mechanism for accounting for variations in products, technologies, and production processes.

Even if a larger number of observations were available, employing data reduction techniques has many benefits. First, reduced-form analysis will always increase the number of degrees of freedom available for hypothesis testing. Second, principal component and factor analyses provide new insights into the underlying regularities of the data. For instance, results from both principal component analysis and factor analysis suggest components and factors that

[19] Additionally, the study has recorded responses to literally hundreds of questions ranging from human resource policies to information processing policies with the intent of identifying practices leading to high performance.

are intuitively appealing and resonate with important aspects of competition within the semiconductor industry. While interpreting principal components and factors in general can be difficult, the techniques offer advantages over less rigorous approaches. Simple plots and charts of performance metrics, for instance, were first used to compare the fabs. But drawing conclusion from these charts was not only difficult but may have lead to incorrect assessments.

The empirical results of the semiconductor data provide a case in point. Principal component analysis reveals that low cycle time co-varies with high labor productivity, high line yields, and high stepper throughput resulting in eigenvector M_PRIN1 (efficient responsiveness). Also, low defect densities co-vary with high cycle times and low engineering effort resulting in eigenvector M_PRIN2 (mass production). These orthogonal vectors were not apparent in individual plots and charts of the variables. Indeed, the principal components for both means and rates-of-change seem intuitively sensible to an informed observer once the underlying relationships are revealed. A similar assertion can be made for the reduced-form factors.

Third, regression analyses which identify the relationship between reduced-form exogenous variables and reduced-form performance metrics identify correlations that otherwise might not be so easily discernible. The correlation between latent technology structure and firm performance will not necessarily be revealed by alternative formulations. For instance, the lack of observations prohibits regressing the 11 exogenous variables onto each of the 14 summary performance statistics. Furthermore, interpreting and summarizing the relationship between right-hand and left-hand variables is more difficult for eleven variables than for five.

When employing the proposed methodology, several caveats must be kept in mind. Many researchers reject the use of exploratory factor analysis because of its atheoretical nature (principal component analysis is less problematic because it produces an orthogonal transformation). We note, however, that factor analysis is used to, in essence, generate proxies instead of directly testing hypotheses. Nevertheless, the fact that factors are not unique suggests that any particular latent structure may not have a relevant physical interpretation and thus may not be suitable for hypothesis testing. Correspondingly, interpreting the physical significance of

particular principal components and factors poses a challenge. While a precise understanding of components and factors is available by studying the loadings, applying a label to a component or factor is subjective and researchers may differ in the labels they use. Yet finding an appropriate label is useful because it facilitates interpretation of regression results and limits the need to work backwards from regression results to component and factor loadings. Nonetheless, the subjectiveness of labels is problematic. Because interpretation of factor loadings is subjective, we recommend that the results of factor analysis be evaluated for relevancy by industry experts before using it in a regression analysis. Also, the robustness of our methodology has yet to be determined. As discussed in Section 3.2, exploratory factor analysis may lack sufficient robustness to be applied in situations when data is non-normally distributed.

Another criticism is that data reduction techniques reduce the richness and quality of the data and thus reduce and confound the data's information content. Data reduction is accomplished by throwing away some data. While throwing away data seems anathema to most practitioners and researchers (especially after the cost incurred for collecting data), principal component analysis and factor analysis retain data that explain much of the variance and omit data that explain little of the variance. Thus, it is unlikely that the application of data reduction techniques will lead to the omission of key information. Obviously, collecting more data and improving survey design is one way to obviate the need for data reduction. Unfortunately, data collection involving large numbers of observations often is impossible either because of a small number of firms or because of the proprietary nature of much of the data. Theoretically, improving survey design could mitigate the need for some data reduction by improving the nature of the data collected. The authors have found, however, that the multidisciplinary nature of the groups engaged in benchmarking efforts coupled with budget and time constraints for designing and implementing surveys invariably leads to tradeoffs that preclude design and implementation of a perfect study. As with all empirical studies, our methodology attempts to make the most out of the data available.

Accounting for non-contemporaneous measurements in the regression analysis rather than in the data reduction step may lead to biases. Analysis of industries with high rates-of-

change, such as in semiconductor fabrication, or where time between observations is large should proceed with caution. A further problem with the method is that even though the degrees of freedom are more likely to be positive after data reduction techniques are applied, six degrees of freedom as in the case of this preliminary study offers a very small number with which to test hypotheses and, thus, is problematic.

The methodology also poses problems for practitioners. The methodology is data intensive, which poses data collection problems. Also, the observation is omitted if any data is missing. If data collection hurdles can be overcome, many practitioners may not be familiar with the statistical concepts employed or have access to the necessary software tools. Both problems can be overcome by collaborative efforts between practitioners (who have access to data) and researchers (who are familiar with statistical techniques and have access to the necessary software tools). Indeed, these reasons resonate with the motivation behind the Alfred P. Sloan Foundation's series of industry studies. These caveats notwithstanding, the proposed methodology offers an exciting opportunity to introduce more systematic analysis and hypothesis testing into benchmarking studies.

Our approach also offers several opportunities for future research. One opportunity is to collect data on additional fabs and expand our analysis. At present, we have incomplete data on several fabs. Filling in the incomplete data would expand our sample and allow us to test our hypotheses with greater precision. Moreover, the data set is likely to grow because CSMS continues to collect data in fabs not in our data set. Perhaps the greatest opportunity to use this methodology is in conjunction with exploring the influence of organizational practices on performance. Organizational hypotheses concerning what forms of organization lead to performance improvement can be developed and tested. CSMS collected data on a large number of variables. These data can be reduced and analyzed in much the same way as the technology metrics. For example, the latent structure of a group of variables describing certain employment practices such as teams and training could be identified via factor analysis and included in the regression analysis.

6. CONCLUSION

Systematically linking performance to underlying practices is one of the greatest challenges facing benchmarking efforts. With the number of observed variables often numbering in the hundreds, data analysis has proven problematic. Systematic data analysis that facilitates the application of hypothesis testing also has been elusive.

This paper proposed a new methodology for resolving these data analysis issues. The methodology is based on the multivariate data reduction techniques of principal component analysis and exploratory factor analysis. The methodology proposed undertaking principal component analysis of performance metrics' summary statistics to construct a reduced-form performance vector. Similarly, the methodology proposed undertaking exploratory factor analysis of independent variables to create a reduced-form set of decision variables. Data reduction increases the degrees of freedom available for regression analysis.

By empirically testing the methodology with data collected by the Competitive Semiconductor Manufacturing Study, we showed that the methodology not only reveals underlying empirical regularities but also facilitates hypothesis testing. Regression analysis showed that while product, technology, and production process variables greatly influence the reduced-form mean performance metrics, they had little impact on the reduced-form rate-of-change performance metrics. Importantly, the proposed model presents a baseline for jointly examining other hypotheses about practices that lead to high performance. Perhaps with the application of the proposed model, practitioners and researchers can employ more systematic analysis to test hypotheses about what really drives high performance.

ACKNOWLEDGMENTS

This work was supported in part by the Alfred P. Sloan Foundation grant for the study on "Competitive Semiconductor Manufacturing" (CSM). We would like to thank all the members of the CSM study at U.C. Berkeley, especially David Hodges, Robert Leachman, David Mowery, and J. George Shanthikumar, for their encouragement and support. Also, we wish to thank three anonymous reviewers whose comments lead us to greatly improve this paper.

REFERENCES

AHIRE, S.L., GOLHAR, D.Y., and WALLER, M.A., 1996, Development and validation of TQM implementation constructs. *Decision Sciences*, **27**, 23–56.

CHARNES, A., COOPER, W.W., and RHODES, E., 1978, Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**, 429–444.

COLLIER, D.A., 1995, Modelling the relationships between process quality errors and overall service process performance. *International Journal of Service Industry Management*, **6**, 4–19.

DILLON, W.R., and GOLDSTEIN, M., 1984, *Multivariate Analysis: Methods and Applications*, (New York: John Wiley & Sons).

FAWCETT, S.E., and CLOSS, D.J., 1993, Coordinated global manufacturing, the logistics/manufacturing interaction, and firm performance. *Journal of Business Logistics*, **14**, 1–25.

HENDRICKS, K.B., and SINGHAL, V.R., 1996, Firm characteristics, total quality management, and financial performance: An empirical investigation. In *Proceedings of the 1996 MSOM Conference*, Institute for Operations Research and the Management Sciences, 128–133.

INDUSTRIAL PERFORMANCE CENTER, 1993, Proceedings of the second Sloan industry studies workshop, Massachusetts Institute of Technology, USA.

JAYANTHI, S., KOCHA, B., and SINHA, K.K., 1996, Measures and models of operational competitiveness: An application to plants in the food processing industry. In *Proceedings of the 1996 MSOM Conference*, Institute for Operations Research and the Management Sciences, 207–212.

JUDGE, G.G., HILL, R.C. GRIFFITHS, W.E., LUTKEPOHL, H., and LEE, T.-C., 1988, *Introduction to the Theory and Practice of Econometrics*, (New York: John Wiley & Sons).

LEACHMAN, R.C. (ed.), 1994, Second report on results of the main phase, Competitive Semiconductor Manufacturing Program Report CSM-08, University of California, Berkeley, USA.

LEACHMAN, R.C., and HODGES, D.A., 1996, Benchmarking semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, **9**, 158–169.

MCNAIR, C.J., and LEIBFRIED, K., 1992, *Benchmarking: A Tool for Continuous Improvement*, (New York: Harper Collins).

POWELL, T.C., 1995, Total quality management as a competitive advantage: A review and empirical study. *Strategic Management Journal*, **16**, 15–37.

SAKAKIBARA, S., FLYNN, B.B., SCHROEDER, R.G., and MORRIS, W.T., 1996, The impact of just-in-time manufacturing and its infrastructure on manufacturing performance. *Management Science*, **43**, 1246–1257.

SZULANSKI, G., 1996, Exploring internal stickiness: Impediments to the transfer of best practices within the firm. *Strategic Management Journal*, **17**, 27–43.

WOMACK, J.P., JONES, D.P., and ROOS, D., 1990, *The Machine that Changed the World*, (New York: Rawson Associates).

Table 1A: Means of Performance Metrics

FAB	CTPL	DLP	ENG	LYD	TLP	STTP	YDD
1	3.596	16.894	81.164	92.863	10.326	232.101	0.970
2	1.583	29.357	352.688	92.001	16.190	318.373	15.194
3	3.150	32.708	121.690	95.952	19.276	319.322	0.754
4	3.311	15.642	167.355	86.766	11.592	328.249	0.419
5	2.611	32.310	87.993	90.152	20.177	491.632	0.491
6	2.489	5.734	24.815	80.402	3.404	143.912	0.431
7	3.205	7.924	27.645	88.438	2.613	221.676	0.846
8	2.734	9.612	25.017	90.501	4.253	13.825	0.990
9	2.901	22.621	95.331	98.267	13.408	379.470	0.290
10	2.002	63.551	205.459	98.460	37.759	606.147	0.313
11	2.291	25.465	100.685	94.543	13.701	259.585	1.895
12	2.711	18.324	91.268	93.484	10.299	203.731	2.476
Mean	2.720	23.350	115.090	91.820	13.580	293.170	2.090
Std. Dev.	0.570	15.630	92.620	5.100	9.550	155.390	4.180

Table 1B: Pearson Correlation for Means of Performance Variables*

	CTPL	DLP	ENG	LYD	TLP	STTP
DLP	-0.481					
ENG	-0.595	0.564				
LYD	-0.128	0.671	0.319			
TLP	-0.427	0.992	0.562	0.634		
STTP	-0.275	0.852	0.498	0.488	0.882	
YDD	-0.624	0.088	0.781	0.035	0.045	-0.015

* Correlations whose absolute value are greater than 0.172 are significant at the 0.05 level: N = 12.

Table2A: Average Rates-of-Change of Performance Metrics

FAB	R_CTPL	R_DLP	R_ENG	R_LYD	R_TLP	R_STTP	R_YDD
1	0.016	0.066	0.067	0.004	0.059	0.108	0.013
2	-0.087	0.095	0.059	-0.001	0.079	0.138	0.020
3	0.005	-0.047	-0.113	0.002	-0.053	-0.030	-0.031
4	-0.018	0.021	0.178	-0.003	0.060	0.047	-0.086
5	-0.083	0.704	0.723	0.008	0.697	0.580	-0.047
6	0.040	0.012	0.030	0.091	0.012	0.017	-0.065
7	-0.135	0.173	0.222	0.022	0.227	0.115	-0.440
8	0.032	0.038	0.076	0.005	0.064	0.271	-0.159
9	-0.004	0.036	0.093	0.004	0.063	0.054	-0.038
10	-0.039	0.045	0.055	0.001	0.042	-0.002	-0.021
11	-0.097	0.018	-0.023	0.005	0.016	-0.015	-0.094
12	0.002	0.016	-0.042	0.007	0.002	-0.004	-0.090
Mean	-0.031	0.098	0.110	0.012	0.106	0.107	-0.087
Std. Dev.	0.057	0.198	0.213	0.026	0.198	0.172	0.122

Table 2B: Pearson Correlation Analysis for Rates-of-Change of Performance Variables*

	R_CTPL	R_DLP	R_ENG	R_LYD	R_TLP	R_STTP
R_DLP	-0.448					
R_ENG	-0.406	0.958				
R_LYD	0.264	-0.050	-0.042			
R_TLP	-0.468	0.993	0.977	-0.050		
R_STTP	-0.233	0.903	0.889	-0.102	0.905	
R_YDD	0.439	-0.065	-0.139	-0.156	-0.151	-0.045

* Correlations whose absolute value are greater than 0.172 are significant at the 0.05 level: N = 12.

Table 3A: Technology Metrics

FAB	STARTS	W_SIZE	FLOWS	P_TYPE	D_TYPE	TECH	P_AGE	D_SIZE	F_SIZE	CLASS	F_AGE
1	2728	6	4	0.5	50	1.1	27	0.73	0	2	0
2	11027	4	4	0	180	2	45	0.03	0	2	-1
3	14467	6	94	0.5	400	0.7	24	0.83	1	2	0
4	5532	5	12	0	200	0.9	36	1.61	1	3	-1
5	6268	6	5	0.5	40	0.8	3	0.42	0	3	1
6	1705	6	7	0	600	0.7	15	1.40	-1	2	-1
7	700	6	1	0	13	0.7	24	1.91	0	1	0
8	350	6	2	0	10	1	12	0.80	-1	2	-1
9	3019	6	5	0	85	0.7	7	0.76	0	1	0
10	6232	6	3	0	15	0.6	9	0.69	1	1	1
11	2172	6	9	0	20	0.8	30	0.42	-1	0	0
12	3453	5	10	0	400	1.2	9	0.36	-1	2	0
Mean	4804	5.7	13.0	0.1	168	0.9	20	0.83	-0.1	1.8	-0.2
Std. Dev.	4257	0.7	25.7	0.2	197	0.4	13	0.55	0.8	0.9	0.7

Table 3B: Pearson Correlation Analysis for Technology Metrics*

	STARTS	W_SIZE	FLOWS	P_TYPE	D_TYPE	TECH	P_AGE	D_SIZE	F_SIZE	CLASS
W_SIZE	-0.127									
FLOWS	0.619	0.050								
P_TYPE	0.420	0.333	0.225							
D_TYPE	0.251	-0.091	0.405	-0.092						
TECH	-0.025	-0.907	-0.093	-0.261	-0.080					
P_AGE	0.237	-0.330	0.140	-0.170	-0.083	0.270				
D_SIZE	-0.212	0.482	-0.088	-0.059	0.125	-0.583	0.087			
F_SIZE	0.571	0.112	0.415	0.407	-0.173	-0.277	0.263	0.146		
CLASS	0.195	-0.406	0.053	0.300	0.285	0.339	-0.302	-0.115	0.000	
F_AGE	0.197	0.570	-0.034	0.365	-0.277	-0.552	-0.441	-0.003	0.239	-0.232

* Correlations whose absolute value are greater than 0.172 are significant at the 0.05 level: N = 12.

Table 4A: Principal Component Loadings for Means of Performance Variables

	M_PRIN1	M_PRIN2	M_PRIN3	M_PRIN4	M_PRIN5	M_PRIN6	M_PRIN7
CTPL	-0.309	0.421	0.356	0.697	0.206	0.267	-0.026
DLP	0.472	0.203	-0.107	-0.127	0.355	0.359	-0.674
ENG	0.388	-0.382	0.128	0.477	0.290	-0.600	-0.121
LYD	0.327	0.270	0.807	-0.304	-0.230	-0.148	0.040
TLP	0.466	0.231	-0.160	-0.006	0.399	0.164	0.720
STTP	0.415	0.267	-0.311	0.390	-0.712	0.003	-0.026
YDD	0.187	-0.662	0.267	0.161	-0.169	0.625	0.103
Eigenvalue	4.007	1.797	0.603	0.441	0.112	0.037	0.002
Proportion	0.572	0.257	0.086	0.063	0.016	0.005	0.000
Cumulative	0.572	0.829	0.915	0.978	0.994	1.000	1.000

Table 4B: Principal Component Loadings for Rates of Performance Variables

	R_PRIN1	R_PRIN2	R_PRIN3	R_PRIN4	R_PRIN5	R_PRIN6	R_PRIN7
R_CTPL	-0.264	0.585	0.305	-0.632	0.258	0.169	-0.003
R_DLP	0.487	0.116	0.041	0.126	0.085	0.616	-0.587
R_ENG	0.482	0.087	0.077	-0.036	0.541	-0.651	-0.191
R_LYD	-0.056	-0.002	0.908	0.389	-0.131	-0.068	0.011
R_TLP	0.493	0.058	0.058	0.034	0.195	0.310	0.785
R_STTP	0.453	0.225	0.044	-0.332	-0.758	-0.238	-0.008
R_YDD	-0.098	0.763	-0.264	0.566	-0.064	-0.103	0.056
Eigenvalue	4.057	1.286	1.120	0.414	0.090	0.032	0.000
Proportion	0.580	0.184	0.160	0.059	0.013	0.005	0.000
Cumulative	0.580	0.763	0.923	0.982	0.995	1.000	1.000

Table 5A: Eigenvalues for Technology Metrics

	1	2	3	4	5	6	7	8	9	10	11
Eigenvalue	3.109	2.470	1.620	1.484	0.946	0.500	0.412	0.212	0.131	0.089	0.028
Proportion	0.283	0.225	0.147	0.135	0.086	0.045	0.037	0.019	0.012	0.008	0.003
Cumulative	0.283	0.507	0.655	0.789	0.875	0.921	0.958	0.978	0.989	0.997	1.000

Table 5B: Loadings for Rotated Technology Factors

	FACTOR1	FACTOR2	FACTOR3	FACTOR4
STARTS	-0.148	0.891	-0.036	0.140
W_SIZE	0.885	0.042	-0.319	-0.171
FLOWS	0.057	0.721	0.095	0.398
P_TYPE	0.061	0.580	-0.558	-0.118
D_TYPE	0.087	0.090	0.005	0.916
TECH	-0.930	-0.156	0.240	0.031
P_AGE	-0.124	0.294	0.868	-0.125
D_SIZE	0.754	-0.106	0.221	0.218
F_SIZE	0.165	0.807	0.098	-0.223
CLASS	-0.489	0.102	-0.455	0.503
F_AGE	0.405	0.230	-0.582	-0.470
Variance Explained by Each Factor	2.700	2.500	1.838	1.648

Table 6: Regression Analysis

	(1) M_PRIN1 (Efficient Responsiveness)	Std. Err.	(2) M_PRIN2 (Mass Production)	Std. Err.	(3) R_PRIN1 (Throughput Improvement)	Std. Err.	(4) R_PRIN2 (Defect Density Improvement)	Std. Err.	(5) R_PRIN3 (Line Yield Improvement)	Std. Err.
Intercept	-1.587	1.279	-0.091	0.639	1.474	1.790	0.735	0.851	0.322	0.550
FACTOR1 (Process Technology Generation)	-1.270	0.690	1.186	0.345 **	0.153	0.966	-0.305	0.460	0.599	0.297 *
FACTOR2 (Process Scale and Scope)	1.622	0.603 **	0.361	0.301	-0.580	0.843	-0.163	0.401	-0.666	0.259 **
FACTOR3 (Process and Facility Age)	0.078	0.503	-0.699	0.251 **	-1.221	0.704	-0.804	0.335 *	-0.264	0.216
FACTOR4 (Product Scope)	-0.641	0.458	-0.309	0.229	-0.751	0.641	0.229	0.305	0.547	0.197 **
TIME	0.546	0.349	-0.031	0.174	-0.420	0.489	-0.226	0.232	-0.159	0.150
Adj. R^2	0.471		0.706		0.000		0.269		0.649	
F (Model)	2.955		6.277 **		0.949		1.811		7.075 **	

** $p \leq .05$
* $p \leq .10$