Estimating professional service productivity: theoretical model, empirical estimates and external validity

John Wacker, James Hershauer, Kenneth D. Walsh, & Chwen Sheu

Published Version Information

Citation: Wacker, J., Hershauer, J., Walsh, K. D., & Sheu, C. (2014). Estimating professional service productivity: Theoretical model, empirical estimates and external validity. International Journal of Production Research, 52(2), 482-495.

Digital Object Identifier (DOI): doi:10.1080/00207543.2013.836611

Publisher's Link: http://www.tandfonline.com/doi/full/10.1080/00207543.2013.836611#.Uyiw5ahdXL8

ESTIMATING PROFESSIONAL SERVICE PRODUCTIVITY: THEORETICAL MODEL, EMPIRICAL ESTIMATES, AND EXTERNAL VALIDITY

1. Introduction

1.1 Professional Service productivity

Service industries are growing more rapidly than manufacturing industries as a share of global income, making their management extremely important for the economic well-being of many countries. Yet, with all the importance of service industries, there is a lack of conceptual understanding of service productivity: exactly which resources make service organizations competitive and how to measure the productivity of these resources (Jääskeläinen and Lönnqvist 2011, Schmenner 2004, Roth and Menor 2003, Yu and Lee 2009, Talluri *et al*. 2012).

In reality, the importance of service productivity carries conceptual as well as empirical support for the financial performance of organizations. At the conceptual level, productivity is important, since the fewer resources a firm uses to produce specified output, the more competitive the organization should be, *ceteris paribus*. Except for qualitative productivity measures, most studies use a surrogate for an organization's output (e.g., value-added). An output is typically measured as revenues, with adjustments for some contingencies. The resources (except perhaps for labor) are also typically measured in monetary costs (van Biema 1997). The difference between the revenues and resource costs is profitability, which is considered the *raison d'etre* for most private organizations. Most private organizations aim for profitability and use productivity as the most important measure of an organization's financial value (Aggarwal 1980, Kendrick 1977, Dean and Kunze 1992, Chang *et al*. 2008, Hopp *et al*. 2009). Smith and Reece (2004) stated that productivity is empirically supported as a surrogate measure for overall service organization performance. In their words "This finding also adds some degree of confidence to those studies that use productivity as a surrogate measure for business performance" (Smith and Reece 2004).

In summary, productivity is both conceptually and empirically important for a service organization's competitiveness and performance. This study proposes a conceptual model that addresses output and resource measuring issues to develop an assessment for the productivity of professional service. Schmenner (1986, 2004) classified all services into four categories: Service factory, Service shop, Mass service, and Professional service (Figure 1). Generally, professional service is associated with a high degree of interaction/variation and a high degree of labor intensity with long throughput time.

(Insert Figure 1 here)

Examples of professional services provided by highly skilled employees are present in many organizations (Anderson 2001, Hopp *et al*. 2009, Jääskeläinen and Lönnqvist 2011, Napoleon and Gaimon 2004, Talluri *et al*. 2012). Consider the highly skilled employees in accounting, where they have multiple outputs of balance sheets and income statements, as well as numerous internal reports. All of these different outputs have high variability in their preparation due to legal requirements and Federal Accounting Standard Boards and additional legal compliance, such as Sarbanes-Oxley. How to measure accounting productivity is important for determining both staffing requirements and costs. Similar considerations are needed for law firms where legal cases call for highly variable labor requirements of case research, preparation of legal documents, trial lawyers, etc. For medical professions, there is a need to determine staffing requirements for improving patient health. In manufacturing firms, design engineers and production control personnel are evaluated on their productivity.

However, in the large construction projects industry, productivity measures are more complex than those of almost any other service industry, due to project size, number of components, and longer time periods for completion and the design uniqueness of each project. In general, each project is unique, since it can range from less than one million dollars to billions of dollars. Additionally, it literally may have a billion design components, it may take many years to complete, and the technology may change during the course of the project. The more complex the project becomes, the more difficult the project is to evaluate which, in turn, causes more assumptions to be necessary for project evaluation. Despite this complexity, the construction engineering design productivity is critical for project cost estimation, project scheduling, internal performance measures and, ultimately, profitability. However, it must be emphasized that although complexity causes many assumptions, the philosophy of science emphasizes that theory is not about differences, but about commonalities and similarities across diverse times and places (Popper, 1957; Bunge, 1967; Hunt, 1991; Wacker, 1998 and numerous others).

In summary, the understanding of service productivity measurement enables improvement in resource usage for the long term survival of organizations. This large project complexity causes many productivity measurement difficulties that are also present in other service industries due to the complexity of service demand (see Smith and Reece 2004). The importance of any methodology depends on when and where it can be applied. The application to real world practices is external validity (for a complete explanation see Calder, Phillips, and Tybout. 1981, 1982; Winer, 1999).This study tests the external validity of the productivity measurement techniques in a real world environment.

1.2 Research objectives and procedures

Assessing the productivity of professional service is a daunting challenge, due to the lack of common and readily available input and output measures (Griliches 1992, Jääskeläinen and Lönnqvist 2011, Keh *et al*. 2006, Hopp *et al*. 2009, Nachum 1999, Schmenner 2004, Talluri *et al*. 2012). There is a paucity of research on the productivity of professional services. Using professional engineering design productivity as an example, this study intends to expand the understanding of this research issue. Specifically, the purpose of this article is to analyze professional service productivity using professional engineering designers for the conceptual and applied productivity methodology. The project was sponsored by the Contract Construction Institute and required that the actual implementation and methodology be applied to participating design engineering organizations. The productivity methodology highlights the interaction of academics with top-level business engineering managers. As a contribution to professional service productivity literature, this article provides pragmatic integration of theory, mathematical modeling, statistical modeling, and estimation and prediction accuracy for the empirical world of professional service management.

This study follows a general procedure that is applicable to studying the productivity of various types of professional service. We begin with an understanding of the current practices and relevant productivity issues based on a literature review and field study (***Step #1***). This field study involved teamwork among three academics and more than ten construction industry executives throughout the entire project. The executives oversaw the process to provide feedback for productivity estimation. Meanwhile, the academic members assured that the model and analyses followed formal strict academic procedures. Additionally, a series of meetings and workshops was held with engineering managers to identify input and output measures. ***Step #2*** develops a conceptual model and a mathematical representation of the problem, followed by ***Step #3***: data collection and model refinement. ***Step #4*** performs statistical analyses to obtain productivity estimates, and ***Step #5*** reviews and determines interpretability of the estimates with the manager feedback. Finally, ***Step #6*** externally validates the conceptual model by comparing the model's productivity estimates with the current projects in the construction industry. The remainder of the paper details the six-step procedure and concludes with the discussion of managerial implications and limitations.

2. Understand the Problem of Productivity Estimate (Step #1)

Typically, every individual type of professional service is unique, requiring an exhaustive literature review. A field study was implemented to more completely understand the current method for productivity estimation to determine input and output measures. This field study involved teamwork with the construction industry executives and several meetings and workshops with the managers. The field study provided a clearer understanding of the current practice, and challenges associated with measuring construction design productivity.

2.1 The nature and concerns of engineering design productivity

Understanding the nature of engineering design in the construction industry is necessary before undertaking a productivity analysis (Brookes 2012, Salter and Torbett 2003). Generally, construction of a major building involves building owners, building designers, and building contractors, along with the contractor's subcontractors (Dubois and Gadde 2002). The building owners decide on their building requirements. Owners then contract with engineering design companies to design the building. Generally owners, along with the designers, hire a construction company to erect the building. The construction company then may hire subcontractors to deliver and build specific components of the building. In short, the owners specify the design company's engineering requirements, the design company develops construction requirements, and the construction company erects the project to the design company's specific building requirements. Clearly, the competence and productivity of design engineers are key factors in delivering the building capabilities required by the owners.

The field study of the academics-executives team also emphasized that engineering design productivity evaluation requires significant human resources, causing any productivity measurement to be subject to extreme scrutiny by both top engineering executives and engineering staff. On the one hand, design engineers are highly skilled professionals and are very sensitive to any method of evaluation. Many design requirements are considered “one-off”, where engineers must use their experience, education, and skills to determine a specific solution to a challenging design problem. The design company faces pressure to avoid the “commoditization” of engineering skills that many engineers perceive as implied by productivity measurement demeaning the design engineering skills.

In addition to making performance evaluations, design engineering organizations must estimate design hours for planning and budgeting. The design engineering management stresses the accuracy of such estimates, since building owners may use the information to negotiate for lower costs of the design services. Specifically, design engineering companies are concerned that their customers (owners) will use available productivity models to drive the design engineering price to a minimum. Consequently, productivity measurement is important for design engineering organizations in order to evaluate performance, plan for capacity, and estimate cost.

The academic-executive team identified several unique features associated with professional service productivity. First, non-tangible outputs are more difficult to measure than physical outputs, due to the vagueness of their nature. Second, without a clear understanding of service output, it is difficult to measure service quality. Third, there is direct communication between the owner (customer) and design engineering (seller) that inherently confounds the labor productivity, due to time spent with the customer. Fourth, the design engineer may be serving several different customers during a specific time period, making labor hours not directly tied to the service output (Johnson and Jones 2004). These four drawbacks have hindered service productivity studies.

The academic research to adequately address those four issues is scant (Brookes, 2012). Some researchers believe that the service productivity concept is intricately confounded with perceived quality, and that profitability should be measured as quality and profitability rather than productivity (Groenroosa and Ojasalob 2004, Talluri *et al*. 2012). Some researchers argue that service productivity output should be measured in overall firm competitiveness (van Bierma 1997). In engineering design, neither of these measures is likely to be useful, since they are too aggregated (Liao et al. 2012).

2.2 Function point method

Although some academics believe that the professional service classification is a task oriented methodology (Hopp *et al*. 2009), this study chose the more established and philosophically conservative approach "function point method" (Bunge 1967, Albrecht and Gaffney 1983). The function point method has been used extensively for many years for measuring productivity in the computer software industry (Mahmood *et al*. 1996, Rothenberger and Hershauer 1999). Namely, this industry uses the various functions of the program, rather than lines of code, for output measures to indicate productivity. The function point method uses the function of the different portions of a computer program. A **function point** is a unit of measurement to express the amount of business functionality an information system provides to a user. The cost (in dollars or hours) of a single unit is calculated from past projects (Mahmood *et al*. 1996). In this study, the function point is the design of the equipment piece to function within the overall project. Therefore, the function point for each piece of equipment has to function within the engineering specifications. In the construction industry, the function point of each design component is how each component interacts with other components within the overall project. If a design component functions, the function point is utilized as the output measure. Therefore, the function point method provides the underlying conceptualization of the measuring output.

Next for inputs, the literature provides some guidance for understanding the complexity of the labor inputs. The important issue is: what is the skill level of design engineers? The literature on service productivity emphasizes the knowledge base as a key factor of the labor input. Roth and Menor (2003) state that “Brain work is extremely complex, requiring employee and firm creativity, innovation, and pioneering approaches to generate new solutions to new problems.” Consequently, it is necessary to choose appropriate skill measures for assessing the labor input. In this study, components are designed by formally trained engineers. Although there are some differences in quality of formal engineering education and experience, these are not critical to include in the labor input measures.

2.3 Current practice

It is important to recognize how engineering services were currently being evaluated for their productivity. Traditionally, engineering design services have used the number of design drawings as the output measure. However, with current information technology, this measure has become problematic for two reasons: one, some of the design items have current stored drawings from previous projects called “go-bys” and two, many of the individual design items are combined with other design items. Both of these cases cause the design time to vary, causing a concern by both building owners and building design firms. A better measure of design hours and the associated productivity is needed to improve project planning and control.

Most of the productivity research for the construction industry has focused on the measurement and improvement of productivity during the construction phase (Thomas and Kramer, 1988). The Construction Industry Institute has ongoing projects that measure the construction phase’s productivity. In contrast, the design engineering phase of construction has not been researched to the same degree as the construction phase has been. The primary reason for this lack of research has been that design engineering is a professional service and has all the measurement difficulties mentioned above. Additionally, similar to other professional services, such as consulting and law firms, the construction industry has large amounts of time spent in interactions between the customer (building owner) and the design engineers. This amount of time is not captured separately from the direct time an engineer spends on the actual specifications, but is instead assigned to the design time for a project. This inclusion is an external productivity issue and this time is included in the total design engineer time spent with the customer (Johnson and Jones, 2004). The clarification of the project specifications is called "scope" in most engineering firms, and is an important and complex issue that deserves a separate research project in the future. Since these hours are generally captured and are tied to specific projects, they are part of this study's productivity measurement.

Finally, just like many professional services, many skills are required to process the output. Since different engineering disciplines (architectural, civil/structural, electrical, instrumentation and control, mechanical, piping and process) have different types of items to design (called design components), the design hours and associated design components are separated by discipline. The basic productivity relationship is between discipline design hours and quantities of the design components. The design components for each discipline are dictated by the customers (building owners), who require specific functionality.

2.4 Determine measurable inputs and outputs

Led by the executives and engineering managers from various disciplines, several workshops were held during this process to review the development of the productivity estimation methodology and to select those design components that are expected to significantly affect discipline design hours. Meanwhile, the academic members assured the legitimacy and accuracy of the methodology and analyses. In general, each workshop had one engineering design discipline for their design components. The results from those workshops were subsequently used by the academic-executive team to develop a conceptual model and a mathematical model.

3. Develop Conceptual Model and Mathematical Model (*Step #2)*

Based on the field study, literature review, and workshops, the academic-executive team developed a conceptual model to aid with the estimate of design hours (Figure 2). The model highlights specific designed components for estimating engineering hours that are the function points. First, the original design component has a conceptual relationship between the design components and the total time the design engineer takes to design these components (in Figure 2, box labeled number 1). These estimated hours represent the hours spent on a specific set of design components. The assumption of time estimates is derived from a work measurement study that the estimates of standard time are not the shortest (optimal) but the time that 95% of the workers can achieve (Myers and Stewart, 2001). Generally, this time is not the total time an engineer spends designing a set of components, since there are other factors that affect those hours. For instance, one factor affecting design hours is the incoming quality of the specifications from the aforementioned project's scope (in Figure 2, incoming quality is the box labeled number 2). The better the clarity of scope, the fewer hours would be spent on the designed components.

(Insert Figure 2 about Here)

There is one very important factor that affects the interpretation of the empirical results: the interaction effects among the sets of design components (in Figure 2, box labeled number 3). The complexity of a set of designed components is determined by how each component is engineered to interact with other designed components. For instance, the design hours for a motor depend on which components must be provided with power. In practice, the more designed components there are that relate to each other, the higher the complexity.

Next, outgoing designs vary in quality, and the more detailed and accurate a design component is, the more design hours it should take (in Figure 2, box labeled number 4). Last, the non-design time is an important factor; these design hours are such things as meetings, site travel time, etc. (in Figure 2, box labeled number 5). Non-design time is included in the total engineering time, since it is assigned to specific building projects.

In short, the design component time is affected not only by its own engineering time, but also by the input and output quality, by other design components, and by non-design time. The relationship between those variables is complex due to the interactions among the design components. To better understand this complexity, a mathematical analysis of these relationships is developed for interpretation of any empirical research that attempts to relate quantities of discipline design components and discipline design hours.

3.1 Formal theoretical model: the mathematical representation

Based on the conceptual model (Figure 2), a mathematical representation is then defined for estimating expected discipline design hours (DH) given some set of components to be designed (DQ). There are numerous technical complexities that arise during estimation of design hours. The most difficult aspect is the interaction between design quantities and the design hours associated with them. The remainder of this section develops a mathematical model to capture this complexity. The first two variables of the model are:

DH_i = Total design hours for discipline *i*. *i* = 1, 2, 3,…*I* (e.g., civil, electrical, mechanical, etc.)

DH_{ij} = Design hours in discipline *i* for designed item/component *j*. *j* = 1, 2, 3,…*J* (e.g., motor, pump, fan, etc.)

The estimated discipline design hours are related to the number of items designed by each discipline. These are defined as discipline designed quantities (DQ), where

$$DH_{ij} = f_1(DQ_{i1}, DQ_{i2}, DQ_{i3}, \ldots DQ_{iJ}) \qquad i = 1,2,3,\ldots I \qquad (1)$$

DQ_{ij} represents the number of item/component *j* designed in engineering discipline *i*. It is assumed that each discipline's design hours are independent of all other disciplines. Further assume that the discipline design hours are a linear function of the design quantities. This assumption means there is no learning curve for designed components. Although it was a concern of the academic team, construction managers stated that each design is relatively unique for each project, so that the linear assumption was preferred. Therefore, for each design quantity, the estimating form would be:

$$DH_{ij} = \beta_{oij} + \beta_{ij} DQ_{ij} + \varepsilon_{ij} \tag{2}$$

where:

β_{0ij} = The design hours required to set up before beginning design on each specific component unit.

β_{ij} = The design hours required to each specific design component *j*, after setting up.

ε_{ij} = the estimated error for the *ij* design hours.

The error term passed the test for normal distribution. Additionally, the individual observation errors passed traditional tests of distribution for outliers' effects on estimated coefficients: DFFITS (influential outliers), DFBETAs (leverage plots), and COVRATIO (observation omission). These statistics were requested and verified by several of the company statisticians. Note that if only one unit is produced, the total discipline design time for a specific design quantity is the summation of the two estimated coefficients ($DH_{ij} = \beta_{0ij} + \beta_{ij}$). For multiple units, the total design hours is a function of the units designed (DQ_{ij}).

The estimation of Equation 2 would be straightforward if each design quantity were independent of each other design quantity. Unfortunately, the statistical independence of design quantities is usually considered a heroic assumption for engineering designs in the construction industry, since design complexities usually arise from the interaction between design components. Consequently, interaction effects between the design components are not readily estimated by Equation 2. More unfortunate is the fact that the interaction effects may not be first order. For example, a motor may have a fan and a drive generator. In this case, it would be a two-way interaction or a second order interaction among the three components. In reality, the complexities are much greater, since there are many high-order interactions among components. For mathematical example here, a first order interaction will provide a conceptual understanding of the pragmatic difficulties of interactions.

$$DH_{ij} = \beta_{oij} + \beta_{ij} DQ_{ij} + \sum_{k \neq j}^{J} \beta_{0ijk} + \sum_{k \neq j}^{J} \beta_{ijk} DQ_{ijk} \tag{3}$$

where β_{0ijk} is the fixed amount of time to prepare if items *j* and *k* are both present. (This may be thought of as a communication time if the two design quantities are not designed by the same discipline engineer.) β_{ijk} is the estimated interaction effect between designed items *j* and *k*. It represents the additional time it takes if both designed items (*j* and *k*) are present. DQ_{ijk} is the number of units of items/components (*j* and *k*) that must be designed simultaneously for functionality. For instance, if *j* = motor and *k* = pump, then β_{ijk} is the extra amount of time it takes to design the motor given a particular type of pump is also designed. DQ_{ijk} is the number of units of design that incorporate both items. Consequently, the interpretation of the discipline design hours associated with each design quantity is confounded by other design quantities, due to the interaction effects.

Based on Equation (3), the total discipline design hours (DH_i) can be expressed as follows.

$$DH_i = \sum_{j=1}^{J} DH_{ij} \tag{4}$$

This equation is expressed more completely as:

$$DH_i = \beta_{oi} + \sum_{j=1}^{J} (\beta_{oij} + \beta_{ij} DQ_{ij}) + \sum_{k \neq j}^{J} \beta_{0ijk} + \sum_{k \neq j}^{J} \beta_{ijk} DQ_{ijk} \tag{5}$$

where β_{oi} represents the overall setup time for a discipline that is independent of any design quantity. The last two expressions are the same for the design hours as:

$$DH_i = \beta_{oi} + \sum_{j=1}^{J} \beta_{oij} + \sum_{j=1}^{J} \beta_{ij} DQ_{ij} + \sum_{k \neq j}^{J} \beta_{0ijk} + \sum_{k \neq j}^{J} \beta_{ijk} DQ_{ijk} \tag{6}$$

The first, second, and fourth terms are all constants. Consequently, any estimated intercept term would include all 3 constant terms and would not permit separation for interpretation. In short, it is most likely that the estimated intercept term would be larger than the overall intercept term (β_{oi}) for the entire discipline, since it would include both the individual estimate design quantities' constant, plus the interactions effects' constant.

In conclusion, there are complex interactions between designed components that indicate how to interpret the estimates. Since the number of interaction terms is usually extremely large, multiple correlations among the variables determine which variables best capture the engineering design hours. As a result, the estimated coefficients include the correlations with related designed components. Fortunately, OLS multiple regression accounts for correlations among designed components to isolate each designed component's primary effect.

3.2 Problems and difficulties with the current method of using "average design hours"

Typically, engineers use the past average hours per design component to determine expected design hours. In this study, the academic-executive team discovered that this approach causes a severe misestimate of the design hours needed for a discipline, in addition to the practical concern that construction industry firms do not separately gather this information. The following is a mathematical representation that illustrates how the average hours estimate leads to a misrepresentation of the actual productivity. Since each designed quantity may have an interaction with other designed quantities, the relationship between designs would be:

$$\frac{DH_{ij}}{DQ_{ij}} = (\beta_{ij} + \sum_{k \neq j}^{J} \beta_{ijk} DQ_{ijk}) + \frac{\beta_{oij}}{DQ_{ij}} \tag{7}$$

Averaging design hours for each specific item will double count the interaction effects between the designed quantities. For example, suppose a pump (say component j=1) and motor (say component j=2) were designed. The pump would have the β_{i1} design hours (assuming no setup time) and the motor would have β_{i2} design hours (also assuming no setup time). However, if the interaction time of how the pump interacts with the motor is estimated separately, the pump would have the additional time β_{i12} for the interaction with the motor.

For productivity measures, traditional work study methods are applicable and productivity is measured as expected inputs over expected outputs. Both the inputs and outputs are in discipline design hours. The expected hours are computed as expected from output hours (DH_i) and the expected inputs are actual hours performed on each project (AH_i). Mathematically, this productivity index is expressed as:

$$P_i = \frac{AH_i}{DH_i} \tag{8}$$

where AH_i = Actual discipline design hours for discipline i, i = 1,2,3,…I and DH_i is derived from Equation (6).

In summary, the above model provides the underlying criteria for estimating productivity in a service environment where there are interactions between output measures.

4. Data Acquisition and Productivity Estimation: Step #3 & Step #4

Data were collected with the support of the Construction Industry Institute. The academic team-executives followed a formal procedure to ensure that the list of variables created for analysis was both academically sound and pragmatically manageable. Initially, bi-variate estimates with the components and the design hours gave a list of possible variables for analysis. The managers/engineers reviewed that list and suggested additional variables. The statistical analysis was performed and subsequently the list was again reviewed and adjusted by both the academic team and the executives/managers. The objective of this exercise is to ensure a manageable list of variables that is both practically important and statistically significant.

The data collection was extremely expensive, since it required counting every component from the specification sheets and finding the actual hours spent by each discipline. These difficulties are expected in most of the professional services where output is very complex with a high degree of variation. The data set went through verifications by two different academics and between three and six industry members. Originally, there were 120 projects, but one international project was not completed and therefore was omitted from further analysis. The final data set contained 119 projects representing almost $15 billion of construction put in place. Only 19 projects involve all disciplines. The discipline sample size is presented in Table 1 and in Table 3.

It is important to note that the project must be completed before the data may be analyzed. All engineering hours are assigned to specific projects. These hours include requests for information (RFI) that represent all conversations on engineering changes while the project is being completed. Projects are completed on average in about three years (3.06 years), with the longest being almost 7 years (6.76 years), and the shortest being just 8 months. Consequently, statistical techniques are cross sectional even though the data are accumulated over the time from the beginning of the project until the project is completed. The engineers also indicated that some of the original components were redesigned due to technological improvements, but the redesigned components does not significantly affect the design hours needed. Additionally, since the function point model is a statistical estimate, it needed to have a large enough sample to be statistically significant. As a result, any modifications of the estimates would have to be handled as a ‘one-off’ by the engineer managers’ judgment.

There were very few instances of two different design components being designed together multiple times within a project or among multiple projects. Thus, statistical significance is a limited due to small sample size. Consequently, the interaction effect was presented to the managers so that they became aware of this important challenge. Upon their advice, the statistical analysis was performed without including the interaction effect. It should also be emphasized that the estimates with the interaction presented the commonly- known problem of interaction effects of multi-collinearity, whose correction would have required centering data, further complicating the interpretation of the estimated coefficients (Freund and Wilson, 1998).

(Insert Table 1 about Here)

4.1 Statistical technique and analysis

Studies in productivity measurement can be broadly categorized as three diverse techniques: index measurement, linear programming and econometric models (Oum *et al*., 1992; Singh *et al*., 2000). This study applies the statistical (econometric) method with the function point technique to explain the productivity theory.

A series of OLS regression analyses were performed for each design discipline, in order to estimate how many hours each design discipline needed. As an example, Table 2 displays the results for one particular design discipline: civil/structural engineers. Design hours are estimated based on several independent variables including building area, structural concrete area, number of deep foundations, and the amount of steel used. The pragmatic usefulness of the estimates presented a challenge for the managers and engineers regarding the interpretation. For instance, one of the engineers commented that the interpretation of the intercept was problematic. He said, “*I am not going to the Vice President and tell him that we need 3,544 hours before a single component is designed*.” In this case, the interpretation of the intercept is that, if there are no design components then the design hours should be equal to the average design hours. The assumption of having design hours when there are no designed components is illogical and impractical. Therefore, the intercept term was excluded to truly represent a real world scenario. After an extensive discussion of this issue, homogenous regression was applied to avoid the classical management criticism that the theory fails to address practicality (Shubik, 1987). These results of homogenous regressions are presented in Table 3.

Additional analyses were performed for all regressions for outlier analyses, using the DFFITS for influence of a single data point. The DFFITS were all within traditional limits and so DFBETAS were not needed to estimate the influence of a single datum on specific estimated coefficients. The COVRATIO were large, indicating a good degree of precision for the estimated coefficients (Freund and Wilson, 1998 pp.119-144). As a side note, statisticians from one company wanted to know how individual data points affected the estimates, so an explanation of the outliers' effects was given for that company's projects.

<Insert Tables 2 & 3 here>

4.2 Statistical results for all disciplines

This section summarizes the results and discusses pragmatic difficulties with interpreting the results from Table 3. It is important to note that, regardless of estimating technique, all construction projects are unique. The underlying principle is, therefore, to capture the similarities among projects so that each discipline's design hours are representative for most of the projects.

The first impression of the overall results is high degrees of variation, as evidenced by high standard deviations of the estimate for all disciplines on Table 1. This observation confirms that estimating professional services and, in this case, engineering design productivity is highly varied from project to project as evidenced by a large standard deviation. This observation also reinforces the challenges of engineering design productivity estimates. First, similar to other professional services, engineering design can be a creative activity and, therefore, has wide variations due to the degree of creativity of individual engineers (Roth and Menor 2004). Second, the projects vary widely as to their purpose, type, and industry. These variations are not only among engineering organizations but also even within individual engineering organizations, since most organizations design buildings for different customers. Thus, there are inherent building purpose variations that cannot be removed and are considered structurally necessary. Third, some of the non-engineering time is devoted to meetings and interactions with other disciplines that can vary widely by engineering discipline. Fourth, the incoming project requirements vary widely, due to the aforementioned "scope" of the project, so two otherwise similar projects can require substantially different design hours. Yet, as stated earlier, the underlying theory is supported by the statistical significance since theory is about similarities and not about differences, since every instance is unique in time and place (Popper, 1957; Bunge, 1967; Hunt, 1991; Wacker, 1998).

The academic-executive team examined the practicality of the results carefully. For some disciplines, such as architecture and civil engineering, the surrogates were well-received by the managers, since the design quantities (components) made intuitive sense. Naturally, there were numerous designed quantities that were *a priori* believed to be statistically significant from bi-variate correlations but did not enter into the estimates. The managers/engineers were encouraged to add more variables believed to be critical, and the final list of variables is both practically important and statistically significant. Upon review by the industry executives, they agreed that the final surrogates provided a better measure.

In the instrument and control, and process disciplines, the surrogates made intuitive sense since the design quantities were primarily tied to the specific discipline. In these cases, there was minimal discussion on the surrogate measures and they were accepted as providing statistical as well as substantive significance.

There were more discussions and questions regarding the results of electrical and piping disciplines, since several designed quantities (components) were statistically insignificant in the estimate due to their correlation with other designed quantities. This correlation is evidenced by the variance inflation factors (VIF) in Table 3. In the electrical discipline, the insignificant variable was motors/generators and in the piping discipline, the insignificant variable was pipe fittings. These conceptual equations were re-estimated including the high tolerance insignificant designed quantities upon insistence of the engineering executives.

An important lesson learned here is that all surrogates must have substantive significance even though these surrogates did not have statistical significance (for a discussion of the difference see McClosky and Zilack 1996). However, the mechanical discipline required several days of meetings and extensive discussion. Initially, the academic-executive team did not recognize the wide variety of equipment designed as being a problem to the estimate of productivity. Yet, the initial estimate indicated that the statistically significant designed quantities were very dissimilar for each construction project and its associated engineering disciplines. The academic-executive team decided to change how the design quantities are defined for output measure. There is an enormous number of different types of mechanical components used in a given construction project, and very few projects have all the components or even some of the design components. As a result, the construction engineer executives suggested the design quantities be grouped by their general design effort (high, low).

In summary, the results presented in Table 3 illustrate that there are substantive surrogate measures that can provide useful and important substitutes for approximations to estimate productivity of design engineers.

5. Discussion and Verification of the Model (Step #5)

Since this study is to be used to estimate how many hours each design discipline needed, it was important to explain to the managers how each estimate should be interpreted and used. Each estimating equation has a standard error of the estimate that represents the amount of variation that can be expected in a project. These errors are relatively large compared to the mean average of the actual design hours. For example, for the architectural design hours, the standard error of the estimate is 2,702 hours and the mean of the architectural design hours is 8,287. This result means that 65% of projects should fall within a range of (8287+/-2702 hours), or between 5,585 and 10,989 hours from Table 1. A first reaction to this finding may be: what is the use of that estimate? Actually, the raw data standard deviation (calculated from the actual data) is 15,379 or more than five times as high as the standard error of estimate, so the estimate is substantively better than the raw data standard deviations. Yet, managers performed some estimation in their organization and felt these results were superior to their very rough original approximations.

Note the standard error of the estimate is only for a single project. The academic-executive team suggests not using this information for single projects for several reasons. First, the standard error is an approximation of the proper method for predicting single discipline hours (the proper method is beyond the scope of the study. See Murphy (1978) for a more complete and exact explanation.) Second, and more important, there is a better approach for using the predictive equations, namely, comparing groups of projects for each discipline. Evaluation of individual projects is not as accurate and using a group of projects will improve the accuracy of the estimated design hours. In other words, a more accurate approach for output evaluation is to average all the design components for the group of projects to derive an estimate of the discipline design hours needed for that group of projects. This approach applies central limit theorem and it permits a calculation of the standard error of average estimates. The standard error of the average estimate is equal to the standard error of the estimate divided by the square root of the number of projects in the sample. Evidently, estimating the professional service hours is preferable to using a group of outputs.

For example, suppose the architectural design manager wants to determine how the total architectural design productivity is performing. If the last 16 projects had, on average, 100,000 square feet, and the average number of architectural engineering design hours on those 16 projects was 4,500 hours, has this group of projects been approached with more or less productive effort? Namely, how many hours are to be expected and were the design hours within normal expectations? First, the expected number of architectural hours for a building of 100,000 square feet can be calculated from the estimated unstandardized coefficients (from Table 3 is 0.03847). Therefore, the expected number of architectural hours should be equal to the estimated coefficient times the number of square feet or (0.03847*100,000) = 3,847 hours. Also from Table 3 (last column), the standard error of the estimate is 2,702 hours. Based on the central limit theorem, the estimated standard error of the estimate for a group of 16 projects is approximately (2702)/(SQRT(16))=675. Since the actual hours are 4,000 and the expected hours are 3,847, the computed Z score is equal to the (actual hours – expected hours)/(standard deviation) or 0.2272 and that is well within the normal expected range. However, if the actual hours had been 5,355 hours, or the computed Z was greater than 1.96, there is a 95% probability that the department used too much time for the 100,000 square foot building. In this hypothetical example, although the architectural engineers used more than the expected hours, they were still well with the normal range of variation. Overall, this analysis provides a legitimate and reliable productivity assessment. The academic-executive team was pleased with the performance of the model and decided to apply the model to actual projects.

6. Implementation and External Validity (Step #6)

The results of this study were reviewed and verified by the Construction Institute members, utilizing their organizational data. These members later successfully applied the model to their actual design engineering organizations. These companies reported that comparing their newly completed projects with the results here identified disciplines where they were over-performing and some disciplines that needed improvement. These results confirmed to the academic-executive team that the proposed model is capable of producing meaningful productivity estimates for design engineering organizations. (Unfortunately, due to the sensitivity of these estimates and contract confidentiality requirements, these results applications could not be included in this paper.) In short, the proposed mathematical model, empirical method and implementable results are all validated for real world applications. The methodology makes good approximations for actual discipline design hours. Overall, the method utilized here fulfilled what many philosophy of science proponents say is the most important criterion for “good” theory: good predictions (Bunge 1965, Wacker 2004).

7. Conclusions

This study uses engineering design as an example to illustrate the conceptual and technical difficulties for developing service productivity estimates. Specifically, professional service productivity measurement has all the measurement challenges of manufacturing industries, with additional problems of knowledge base and customer-service provider interactions. A team of academic and industry executives was formed to review the unique problems of engineering design productivity. With additional feedback from the managers, the team formulated the productivity estimate problems into a mathematical model. The statistical (econometric) technique with the function point method was applied to analyze the professional service productivity problem. In this study, the analysis of professional service productivity measurement is performed from a conceptual and pragmatic perspective, and the results contribute to a better understanding of how to develop surrogates for outputs and inputs for measuring service productivity. A summary of specific contributions of this study are:

- This study mathematically demonstrated that the current methods of productivity measurement in construction industry design are inherently biased. These biases are eliminated in a more sophisticated statistical model. Evidently, the current productivity method needs improvement.
- This study illustrates the value of a formal academic procedure for pragmatic problems familiar to academics, including identification of the requirements, conceptual model development, mathematical model development, empirical model development, empirical results, and external validation of results. At each stage, managers interacted with academics and provided specific inputs and suggestions to facilitate the productivity measurement.
- This study demonstrates that in a very complex environment the academic procedure provides useful methods to improve productivity measures. In the contract construction industry, design services vary by as much as a thousand fold, the outputs may be in the millions, the outputs takes years to complete, and the technology changes during the completion of a project. Even with such complexity, the proposed procedure proves to be useful for managers.
- A useful service productivity measure is only possible with constant examination and modification(s) of the measurement methods by the managers. These modifications are instrumental for the productivity measurements. It should be noted that these modifications are not only to conceptual and mathematical models, but also to statistical techniques.
- The ultimate validation of any technique is the degree to which it is applied, called external validity (Calder, Phillips, and Tybout. 1981, 1982; Winer, 1999). The proposed procedure and model for productivity measurement were tested and their external validity successfully demonstrated.

In conclusion, this study develops a productivity estimation methodology that highlights the interaction of academics with professional business engineering managers. The managers actually accepted the overall approach and the proposed model very well. They all felt the model provides a better estimate than their current approach. There are only a few exceptions where they insisted on adjusting our model based on their experience. In those instances, the academic team made sure the managers understood the implications of deviating from the theories. This study contributes to professional service productivity literature with a pragmatic integration of theory, statistical modeling, and estimation and prediction accuracy, for the empirical world of professional service management. Future studies using other professional services may wish to include a wider variety of outputs to estimate professional service productivity, including such factors as contract scope and interaction with other activities.

References

Abernathy, F.H., Dunlop, J.T., Hammond, J.H., and Weil, D., 1999. *A stitch in time*. New York: Oxford University Press.

Aggarwal, S.C., 1980. A study of productivity measures for improving benefit-cost ratios of operating organizations. *International Journal of Production Research*, 18 (1), 83-103.

Albrecht, A. and Gaffney, J., 1983. Software function, source lines of code, and development effort prediction: A software science validation. *IEEE Transactions on Software Engineering,* 9 (6), 639-648.

Anderson, E.G., 2001. Managing the impact of high market growth and learning on knowledge worker productivity and service quality. *European Journal of Operational Research,* 134 (3), 508-524.

Balk, B.M., 1998. *Industrial price, quantity, and productivity indices: The micro-economic theory and an application*. Boston: Kluwer Academic.

Brookes, N., 2012. What is engineering construction and why is it important? Toward a research agenda. *Construction Management and Economics*, 30 (8), 603-607.

Bunge, M., 1967. *Scientific research I: The search for system*, New York: Springer-Verlag.

Calder, B.J., Phillips, L.W. and Tybout, A.M., 1981. Designing research for application. *Journal of Consumer Research,* 8, 197-207.

Calder, B.J., Phillips, L.W. and Tybout, A.M., 1982. The concept of external validity. *Journal of Consumer Research*, 9, 240-244.

Chang, D.S., Kuo, Y.C. and Chen T.Y., 2008. Productivity measurement of the manufacturing process for outsourcing decisions: the case of a Taiwanese printed circuit board manufacturer. *International Journal of Production Research*, 46 (24), 6981-6995.

Dean, E. R. and Kunze, K., 1992. Productivity measurement in service industries. (Zvi Griliches edition). *Output measurement in the service sectors*. NBER. *Studies in income and wealth*. Chicago, IL: University of Chicago Press.

Dubois, A. and Gadde, L., 2002. The construction industry as a loosely coupled system: implications for productivity and innovation. *Construction Management and Economics*, 20 (7), 621-631.

Greene, W.H., 2004. *Econometric analysis*. (5th ed.). New York: Pearson Education.

Griliches, Z., 1992. Introduction (Zvi Griliches edition). Output measurement in the service sectors, *NBER studies in income and wealth*. Chicago, IL: University of Chicago Press.

Groenroosa, C. and Ojasalob, K., 2004. Toward a conceptualization of the transformation of inputs into economic results in services. *Journal of Business Research,* 57, 414-423

Hopp, W.J., Iravani, S.M.R., and Liu, F., 2009. Managing white-collar work: An operations-oriented survey. *Production and Operations Management,* 18 (1), 1–32.

Jääskeläinen, A. and Lönnqvist, A., 2011. Public service productivity: how to capture outputs? *International Journal of Public Sector Management*, 24 (4), 289-302.

Johnston, R., and Jones, P., 2004. Service productivity: Towards understanding the relationship between operational and customer productivity. *International Journal of Productivity and Performance Management,* 53 (3/4), 201-213.

Keh, H.T., Chu, S., and Xu, J., 2006. Efficiency, effectiveness and productivity of marketing in services. *European Journal of Operational Research,* 170, 265–276.

Kendrick, J.W., 1977. *Understanding productivity,* Baltimore: Johns Hopkins University Press.

Kozicki, S., 1997. The productivity growth slowdown: Diverging trends in the manufacturing and service sectors. *Economic Review -Federal Reserve Bank of Kansas City,* 82 (1), 31-47.

Liao, P., Stephen R., Thomas, W., O'Brien, J., Dai, J., Mulva, S.P., and Kim, I., 2012. Benchmarking project level engineering productivity. *Journal of Civil Engineering and Management*, 18 (2), 235-244.

Mahmood, A.A., Pettingell, K.J., and Shaskevich, A.L., 1996. Measuring productivity of software projects: A data envelopment analysis approach. *Decision Sciences,* 27 (1), 57-80.

McCloskey, D. and Ziliak, S., 1996. The standard error of regression. *Journal of Economic Literature*, 34 (1), 97-114.

Murphy,J.L. 1973. Introductory Econometrics. Richard D. Irwin, Inc. Homewood, IL.

Myers, F.E. and Steward, J.R., 2001. *Time and motion study*. Upper River, NJ: Prentice Hall.

Nachum, L., 1999. Measurement of productivity of professional services: An illustration of Swedish management consulting firms. *International Journal of Operations & Production Management*, 19 (9), 922 - 950.

Napoleon, K. and Gaimon, C., 2004. The creation of output and quality in services: A framework to analyze information technology-worker systems. *Production and Operations Management*, 13 (3), 245-259.

Oum, T. H., Tretheway, M.W., and Waters, W.G.I., 1992. Concepts, methods and purposes of productivity measurement in transportation. *Transportation Research Part A: Policy and Practice,* 26A (6), 493-505.

Roth, A.V. and Menor, L.J., 2003. Insights into service operations management: A research agenda. *Production and Operations Management,* 12 (2), 145-165.

Rothenberger, M.J. and Hershauer, C., 1999. A software reuse measure: Monitoring an enterprise-level model driven development process. *Information & Management*, 35, 283-293.

Salter, A. and Torbett, R., 2003. Innovation and performance in engineering design. *Construction Management and Economics*, 21 (6), 573-580.

Smith, T.M. and Reece, J.S., 2004. The relationship of strategy, fit, productivity, and business performance in a services setting. *Journal of Business Research*, 57, 414-423.

Schmenner, R., 2004. Service businesses and productivity. *Decision Sciences*, 35 (3), 333-347.

Schmenner, R., 1986. How can service businesses prosper? *Sloan Management Review*, 28 (3), 21-32.

Shubik, M., 1987. When is it an application and when is theory a waste of time? *Management Science*, 33 (1), 1511–1522.

Singh, H., Motwani, J., and Kumar, A., 2000. A review and analysis of the state-of-the-art research on productivity measurement. *Industrial Management & Data Systems*, 100 (5), 234

241.

Talluri, S., Kim, M.K., and Schoenherr, T., 2012. The relationship between operating efficiency and service quality: are they compatible? *International Journal of Production Research*, Available online, 30 Oct 2012.

Thomas, H.R. and Kramer, D., 1988. The manual of construction productivity measurement and performance evaluation. *CII Source Document 35*. Austin, TX: Construction Industry Institute.

van Biema, M., 1997. Managing our way to higher service-sector productivity. *Harvard Business Review*, 75 (4), 87-95.

Wacker, J.G., 2004. A theory of formal conceptual definitions: Developing theory-building measurement instruments. *Journal of Operations Management*, 22 (4), 629-650.

Winer, R.S. 1999. Experimentation in the 21st Century: The Importance of External Validity. *Journal of the Academy of Marketing Science,* 27(3), 349-358.

Yu, M. and Lee, B., 2009. Efficiency and effectiveness of service business: Evidence from international tourist hotels in Taiwan. *Tourism Management*, 30 (4), 571-580.

Figure 1. Service matrix (Schmenner, 1986, 2004)

		Degree of Customization & Variation	
		Low	***High***
Labor Intensity & Throughput Time	***High***	**Mass Service** (e.g., Retailing, banking, boutiques)	**Professional Service** (e.g., Law firms, architecture design)
	Low	**Service Factory** (e.g., Trucking, warehouse)	**Service Shop** (e.g., Airline, restaurant)

Figure 2. Conceptual model for estimating engineering hours

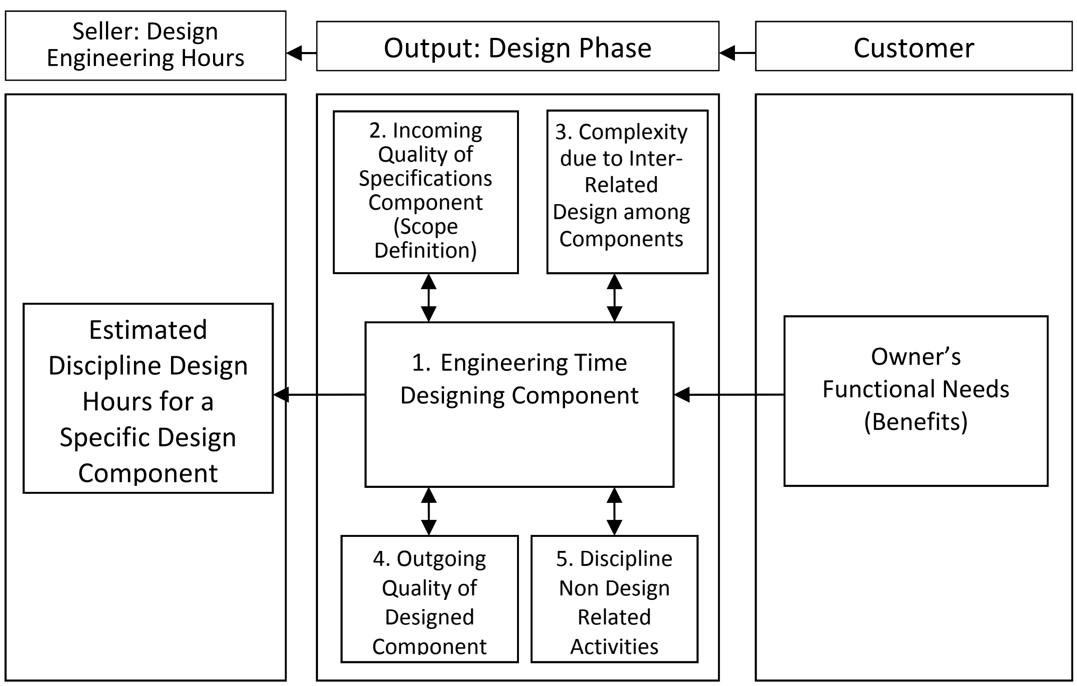

Table 1. Descriptive statistics of disciplines

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Architectural: Actual Hours	25	441	61219	8287	15380
Civil: Actual Hours	105	18	261568	17875	33656
Electrical: Actual Hours	110	30	326272	17995	34249
Instrumentation: Actual Hours	83	31	125280	8792	16988
Mechanical: Actual Hours	104	15	311928	21439	37360
Piping: Actual Hours	79	58	333654	20700	45572
Process: Actual Hours	63	26	36958	5522	9099

Table 2. Civil/Structural engineering hours estimate with intercept

	Coefficients	*Standard Error*	*T Stat*	*One tail P-value*
Intercept	3544.0	1202.4443	2.9474	0.0022
Building area square footage	0.01905	0.0098	1.9412	0.0281
Number of Deep Foundations (piles, piers, caissons)	3.03728	1.6374	1.8549	0.0339
Structural Concrete cubic yards	0.37588	0.1083	3.4697	0.0005
Steel: Tons of structural, pipe rack, utility structural & misc.	5.45504	0.7694	7.0903	0.0000

N=76; R square= 0.6801; Adjusted R square= 0.6621;
Standard error of the estimate (root mean squared error) = 7251.31

Table 3. Overall regression estimates of discipline hours on design quantities

	Unstandardized Coefficients				Collinearity Statistics						Std. Error of the Estimate
Architectural Discipline	B	Std. Error	t	Sig.	**Tolerance**	VIF	Beta	R Square	Adj. R Square	Sig. F Change	Std. Error
Usable Building area	0.03847	0.003	12.622	0.000	1	1	0.948	0.898	0.893	0	2702.7384
	n=25										
Civil Discipline	B	Std. Error	t	Sig.	**Tolerance**	VIF	Beta	R Square	Adj. R Square	Sig. F Change	Std. Error
Useable Building area	0.0161	0.01	1.562	0.123	0.626	1.597	0.094	0.837	0.828	0.00	7628.58765
Number of Deep Foundations	3.53	1.714	2.06	0.043	0.674	1.483	0.119				
Structural Concrete	0.492	0.106	4.63	0	0.469	2.130	0.322				
Steel - Tons of structural,pipe rack, utility structural & misc steel	6.385	0.738	8.648	0	0.558	1.791	0.551				
	n=105										
Electrical Discipline	B	Std. Error	t	Sig.	**Tolerance**	VIF	Beta	R Square	Adj. R Square	Sig. F Change	Std. Error
Number of terminations	0.234	0.115	2.035	0.046	0.124	8.078	0.286	0.853	0.846	0.00	10493.0992
Linear feet of cable	0.0160	0.005	3.556	0.001	0.088	11.328	0.592				
Number of Motors/Generators	7.60	17.504	0.434	0.666	0.117	8.539	0.063				
	n=110										
Instruments and Controls Discipline	B	Std. Error	t	Sig.	**Tolerance**	VIF	Beta	R Square	Adj. R Square	Sig. F Change	Std. Error
Number of tagged devices	5.03	0.699	7.193	0.000	1	1	0.639	0.408	0.4	0.00	14378.664
	n=83										
Mechanical Discipline	B	Std. Error			**Tolerance**	VIF	Beta	R Square	Adj. R Square	Sig. F Change	Std. Error
High effort design components	1295.18	234.313	5.528	0	0.302	3.314	0.497	0.808	0.8	0.00	5962.86849
Low effort design components	50.4	11.744	4.29	0	0.262	3.819	0.414				
Other Equipment components	3.832	3.554	1.078	0.284	0.76	1.315	0.061				
	n=104										
Piping Discipline	B	Std. Error	t	Sig.	**Tolerance**	VIF	Beta	R Square	Adj. R Square	Sig. F Change	Std. Error
Llinear feet of pipe	0.169	0.036	4.65	0.000	0.276	3.627	0.482	0.864	0.855	0.00	10288.3669
Number of pipe fittings	0.521	0.413	1.26	0.214	0.222	4.501	0.146				
Number of pipe supports and hangers	2.302	0.883	2.606	0.012	0.163	6.123	0.351				
	n=79										
Process Discipline	B	Std. Error	t	Sig.	**Tolerance**	VIF	Beta	R Square	Adj. R Square	Sig. F Change	Std. Error
Number of tagged devices	1.756	0.26	6.741	0.000	0.538	1.857	0.610	0.797	0.788	0	2172.48823
Number of selected equipment items	18.091	4.582	3.948	0.000	0.538	1.857	0.358				
	n=63										