Exploratory Framework for Application of Analytics in the Construction Industry

Nader Naderpajouh, M.ASCE[1]; Juyeong Choi[2]; and Makarand Hastak, M.ASCE[3]

Abstract: The complex dynamics inherent to the context of decision-making in the construction industry requires more rigorous application of analytics. However, effective frameworks to facilitate such data-driven decision-making are noticeably lacking in the construction industry. To address this lack, the Purdue Index for Construction (Pi-C) is introduced in this paper as a collaborative effort to facilitate and promote data-driven decision-making in the construction industry. As a preliminary step, a hierarchical definition for health of the construction industry is explored based on the results of a literature review, survey, and interviews. The developed hierarchical definition is then used to propose a framework to benchmark, interpret, and analyze data associated with the status of the health of the industry. The proposed framework is tested with existing publicly-available data to explore its effectiveness in improving decisions made in the form of policies or strategies. The research results highlight the gap in the availability and frequency of data for analytics in the construction industry, the need for benchmarking the dynamics of the industry as a coupled system, and the potential for using analytics. Therefore, topics within the construction industry that require more-rigorous data collection were systematically explored. Policy-makers and strategy developers can apply the proposed framework for data-driven decision-making using their preferred set of data as well as communication of data on trends. Researchers can use this framework to further explore the dynamics of the health of the construction industry on topics such as sustainable development or the diversity of the construction project areas. **DOI: 10.1061/(ASCE)ME.1943-5479.0000409.**

Author keywords: Purdue index for construction; Data analysis; Policies; Data communication; Sustainable development; Economic factors; Social factors.

Introduction

The complexity of the construction industry as a loosely-coupled system (Dubois and Gadde 2002) is apparent in the diversity of the actors and their interactions as well as the diverse specialties involved and the nonrepetitive nature of construction projects (Pries and Janszen 1995; Baccarini 1996; Fernandez Solis 2008). The complexity of the construction industry is coupled with its significance in the global economy. Past studies suggested the importance of the construction industry to economic development, albeit more significant in impact in the developing rather than developed contexts (Turin 1978; Wells 1985; Bon 1992; Raftery et al. 1998; Ruddock and Lopes 2006; Giang and Pheng 2011). In the United States alone, the construction industry accounted for approximately 3.82% of the gross domestic product (GDP) (BEA 2014) and provided 9.27 million jobs in 2013 (BLS 2014a). Regardless of the magnitude of its contribution to the GDP, the importance of the construction industry is highlighted by the vitality of its output to address market dynamics in events such as recessions (Gregori and Pietroforte 2015). Observing the dynamics of this coupled system using a multifaceted approach could address the need of the industry to better understand its complex nature.

Data-driven analytics has gained considerable attention in recent years due to their ability to enable identification of trends and patterns of dynamics for business intelligence. The major goal of this surge is to promote data-driven strategy development and policy-making (i.e., informed decisions based on the analysis of data on trends and patterns of complex dynamics). Data-driven decision-making substantially improves strategies and policies, enables informed decisions, minimizes risks, and reveals hidden valuable insights (Manyika et al. 2011), especially in complex contexts such as the construction industry. However, trend analysis in construction has traditionally focused on financial dynamics and cost trends with special emphasis on project-level analysis.

Several studies have explored the costs, associated variations, and trends in the construction industry (Hwang 2009; Ashuri and Lu 2010; Xu and Moon 2011; Cao et al. 2015) specifically focused on project budget management while numerous other research studies have focused on financial trends (Yee and Cheah 2006; Jung et al. 2012b; Zilke and Taylor 2015; Yoon et al. 2015; Chiang et al. 2015). The focus on financial and cost issues may not be sufficient to develop long-term and comprehensive strategies (as sequences of actions) or policies (as directions of actions). Long-term policies and strategies should include broader issues such as competitiveness and productivity and thereby expand the analysis to the division of labor and specialties as suggested by classical economists; increasing the emphasis on physical capacity investments as suggested by neoclassical economists; and capacity building through education, training, and technological progress (Momaya and Selby 2009; Jung et al. 2012a; Deng et al. 2013; Schwab 2013). A prerequisite to data-driven decision-making is

[1]Visiting Assistant Professor, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907-2051 (corresponding author). E-mail: nnp@purdue.edu

[2]Ph.D. Student, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907-2051. E-mail: choi287@purdue.edu

[3]Professor and Head, Division of Construction Engineering and Management; and Professor, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907-2051. E-mail: hastak@purdue.edu

Note. This manuscript was submitted on April 30, 2015; approved on August 7, 2015; published online on October 13, 2015. Discussion period open until March 13, 2016; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Management in Engineering***, © ASCE, ISSN 0742-597X.

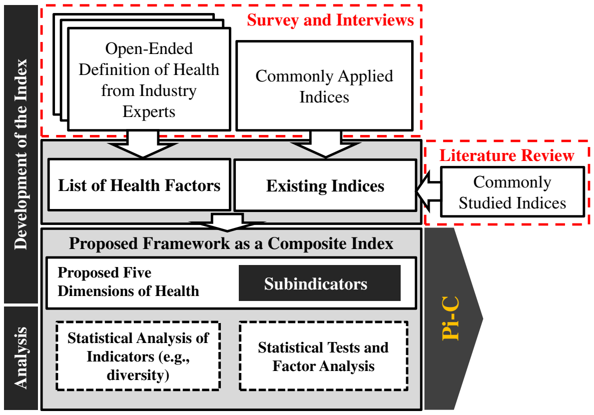

Fig. 1. Research structure

an analysis of the trends for these broader issues beyond financial dynamics and cost trends.

The focus of analysis in the construction industry until now has been limited to mostly project-level data due to the project-driven nature of the industry. Several studies suggested the significant impact of industry factors on the profitability of firms (Schmalensee 1985; McGahan and Porter 1997; Hawawini et al. 2005) while others argued for firm-level metrics as the major determinant of profitability (Rumelt 1991; McNamara et al. 2005; Short et al. 2007; Yoon et al. 2015). Regardless of the contrasting theories, both sides of the argument agreed on the fact that industry-level factors are important to a varying degree to the performance of firms. As a result, firms need to discover the significant factors for their specific case at the firm or industry levels (Phua 2006), as industry-level factors are vital for an average body of firms within the industry whereas firm-related indicators are vital for overperformers or underperformers (Hawawini et al. 2005).

This paper presents research in progress at Purdue University to investigate data-driven policy-making and the use of big data to assess the health of the construction industry. The following section is a discussion of the results of a preliminary survey conducted to understand the status of the industry in the application of indices and the suggestions to establish a definition of health at the industry level collected from the survey of industry experts are discussed in the subsequent section. Then, the proposed Purdue Index for Construction (Pi-C) framework is introduced as a composite indicator and discussed in view of the publically-available data along with associated statistical tests and interpretations. Finally, the practical implications of the Pi-C framework for policy-making in the construction industry are presented. The paper concludes with an outline of the future steps for Pi-C as well as areas that need further data collection within the construction industry. Fig. 1 shows the logical structure of the research and its outcome.

Background

Current Status of Applying Indices

Indices are the medium used to facilitate communication of data associated with trends, trajectories, or patterns of dynamics. For example, the key performance index (KPI) aims to provide a holistic framework to gauge the success of projects (Chan and Chan 2004), while Yu et al. (2007) proposed a model to compare the performance of construction companies. The total recordable incident rate (TRIR) and the lost time incident rate (LTIR) are also common indices for the construction industry to benchmark their safety records or to observe their longitudinal trajectory (Bureau of Labor and Statics 2013). A survey at the initial stages of this research to gain insights into the status of indices in the construction industry indicated that, in addition to the aforementioned indices, the following commonly-applied indices in the construction industry are being used: income per full-time equivalent (FTE), Engineering News Record (ENR) cost indices, construction backlog indicator (CBI), architecture billing index (ABI), construction industry confidence index (CICI), and construction industry round table (CIRT) sentiment index, as well as the trends on prices of oil or other commodities. The survey, reported in Naderpajouh et al. (2012) hinted at the lack of extensive application of indices in the construction industry with a focus on project-level indices as well as the financial aspects of the industry. This gap is underscored by the need for more-effective policies to assist the industry with the increasing requirements of sustainability as well as the increasing challenges for the industry to cope with market volatilities. Although analysis of the performance and financial/cost trends is important to securing short-term achievements, limiting the focus of analytics on these issues may not guarantee sustainable success (De Smet et al. 2007). The sustainable success of a firm requires the application of analytics that include a combination of financial and nonfinancial health indices (De Smet et al. 2007). A framework to address this gap should (1) facilitate a combination of financial and nonfinancial indicators at the industry level; and (2) be flexible enough to address diverse challenges and focus areas within the industry as a coupled system.

Composite Indices

Composite indices aim to address the need for communicating trends that are complex and multifaceted in nature (Booysen 2002; Nardo et al. 2005). A realistic approach to the strengths and weaknesses of these indices is necessary to effectively apply them in construction decision-making. Stiglitz et al. (2010) suggested that the composite indices can help users cope with the abundant information presented in dashboards or sets of indicators. However, the focus of composite indices to cover a broad range of policy issues often discounts meaningful interpretation of the indices (McGillivray and Noorbakhsh 2004). In many cases the composite indices lack a comprehensive meaning and remain as an aggregate of their subindices. Therefore, the composite indicators can be better applied as a navigator through the glut of information in dashboards and may be regarded as the first step to a closer look at the broad spectrum of issues presented through their subindices (Stiglitz et al. 2010). As a result, these indices are most useful for raising the awareness of the industry for data-driven policy-making and strategy development. Effective policies and strategies, however, require further analysis of the components of the index. For example, several composite indices were developed by international organizations such as the United Nations (UN) or the Organization for Economic Cooperation and Development (OECD), which include indicators such as the human development index (Noorbakhsh 1998; UNDP 2007), the ocean health index (Halpern et al. 2012), the global competitiveness index (Schwab 2013), the index for economic well-being (Osberg and Sharpe 2002), the better life index (OECD 2013), the global innovation index (Dutta 2012), the environmental performance index (Esty et al. 2008), and the environmental sustainability index (Esty et al. 2005). These indicators are used to compare different contexts, such as countries or periods of time, in terms of a specific phenomenon, generally of a multifaceted and complex nature such as environmental performance. Although useful for initiating a discussion or supporting a policy, these indices still require a complementary backup of quantitative assessment for data-driven decisions.

Defining the Health of the Construction Industry

As shown in Fig. 1, the development of the proposed index began with a literature review, a survey, and interviews. The survey included an open-ended question to define the health of the construction industry (Naderpajouh et al. 2012). Fowler (1995) argued that the open-ended question encourages genuine responses and eliminates the bias of the research group in the identification of the properties of the phenomenon of health at the industry level. To address the challenge of analyzing the responses, they were coded into individual factors (Table 1) as suggested by Fowler (1995). As a result, a list of mutually-exclusive and collectively-inclusive factors was developed reflecting the survey responses to providing a definition of the health of the industry. This list was then reviewed to ensure inclusion of the measurable factors of health. Subsequently, the list was revised through iterative coding with the aim of achieving a collectively-inclusive and mutually exclusive list of factors. The health factors were ranked based on repetition of each individual factor within the provided responses. Finally, the list of individual factors was discussed and reviewed with selected industry experts for further verification. Table 1 provides the detailed results of the analysis based on the frequency and associated percentage out of 45 responses. The individual factors were then integrated into the proposed five-dimensional framework of health (Naderpajouh et al. 2012), which describes the construction industry as healthy if:

1. The industry indicates positive economic and financial performance (economic dimension);
2. The industry is stable and resilient to internal/external shocks (stability dimension);
3. The industry offers a pleasant working atmosphere for individuals involved within the industry (social dimension);
4. The industry applies the best of the expertise, science, and technology in the production process (development dimension); and
5. The industry produces high-quality products for its users (quality dimension).

The proposed framework to define the health of the construction industry is hierarchical in nature. The economy of the construction industry is the bottom-line of its health, while its stability and resilience to external shocks indicate a higher level of health to ensure the stability of business activity. Both of these dimensions reflect the economic metric of sustainability (Levitt 2007) as well as the first and fourth outcome of the vision for civil engineering in 2025 as it refers to the role of the profession in society's economic engine and management of risk and uncertainty caused by natural events, accidents, and other threats (ASCE 2007). The social health of the industry is a higher-level indicator of its health and reflects the working atmosphere for the workforce. This dimension reflects the social metric of sustainability (Levitt 2007) and the role of the profession in society's social engine (ASCE 2007). One level higher concerns development and indicates the trends of the sustenance of the industry by education and capacity building in human resources, application of innovative technologies, and research. This dimension reflects sustainable development in general (Levitt 2007) as well as innovations and the integration of ideas and technology across the public, private, and academic sectors (ASCE 2007). The highest level of health at the industry level is achieved by the quality of its output, such as projects, plans, or designs. The satisfaction of users with the output enhances the reputation of the industry, and the consideration of environmental impacts ensures the sustenance of the industry. This dimension reflects the environmental metrics of sustainability (Levitt 2007) as well as stewardship of the natural environment and its resources and leadership in shaping public environmental and infrastructure policy (ASCE 2007). The five-dimensional definition of health is the basis of the Pi-C framework for data-driven policy-making in the construction industry. The identified factors, extracted as the coded definition of health from the survey, are integrated within these dimensions.

Table 1. Factors That Define Health of the Construction Industry (Reprinted from Naderpajouh et al. 2012, © ASCE)

Construction health factor	Number	Percentage (%)
Backlog volume	23	53
New investments	17	40
Competitiveness	16	37
Profitability	16	37
Growth trajectory	14	33
Employment/layoff rate	10	23
Diversity of projects	5	12
Academic education/non-academic training	4	9
Market willingness to pay	3	7
Safety	3	7
Average payback period	2	5
Availability of financing	2	5
Quality of work	2	5
Owner's satisfaction	2	5
User's satisfaction	2	5
Business activity in the industry	2	5
Union workers	1	2
Dependencies	1	2
Changes and modifications	1	2
Development	1	2
Contribution to GDP	1	2
Number of disputes	1	2

Pi-C Framework

The Pi-C framework is based on the proposed philosophy of health for the construction industry and its five dimensions: (1) economic (E); (2) stability or resilience (SR); (3) social (S); (4) (sustainable) development (D); and (5) quality (Q). The focus of Pi-C is to gauge the trajectory of the current health of the construction industry. Therefore, the authors refrained from defining the thresholds of health at the construction industry level; Pi-C indicates instead the trajectory of the health status compared to a reference time, chosen as December 2013. Pi-C is a composite index comprised of five dimensions, and each dimension is composed of individual variables that represent the identified factors of health (Table 1) which, together, denote the final index for each dimension. Each variable can be an existing or proposed index by itself. As a result, Pi-C outlines the improvement or deterioration of the health of the industry during each publication period through a cluster of variables.

The mathematical implication of the core idea of Pi-C, which is to observe trajectories without suggesting any thresholds or predetermined benchmarks, is an index in the form of a coefficient of the current status of the variables to the status in the reference time ($t = 0$). This coefficient is established for each variable associated with the health factors. Each dimension is then equal to the geometric average of the corresponding variables. As shown in Eq. (1), each dimension (DI) is defined based on i ($i = 1$ to n) associated variables ($I_{i,t}$) at current time (t) with the option to integrate the correspondent weights (ω_i)

$$\mathrm{DI} = \left(100 \prod_{i=1}^{n} \omega_i \frac{I_{i,t}}{I_{i,t=0}}\right)^{1/n} \tag{1}$$

This approach not only bypasses establishing thresholds but also normalizes different variables of Pi-C through distance to a reference. Normalization is required in composite indices before aggregation (Nardo et al. 2005). Each dimension of Pi-C is then a ratio indicating the downward or upward trajectory of the health from the reference point based on the selected variables. The Pi-C index (π_c) is the geometric average of all the dimensions Eq. (2). The geometric average was selected due to the reduction of compensability among the dimensions (Nardo et al. 2005). The trajectory of each dimension by itself clearly may be beneficial for policy-makers and strategy developers to trace the roots of dynamics

$$\pi_c = (\mathrm{E} \times \mathrm{SR} \times \mathrm{S} \times \mathrm{D} \times \mathrm{Q})^{1/5} \tag{2}$$

A Pi-C value greater than 100 will indicate improvement of the health of the industry compared to the reference point value as a baseline, while a Pi-C less than 100 will signal a deterioration trend in the health of the industry. Although Pi-C theoretically can have any value between 0 (demise of the industry, $t \to E$) and infinity (boundless success of the industry, $t \to P$), $\pi_c \in [0, \infty)$, practically, the upper limit might be bounded

$$\pi_c > 100; \qquad -\lim_{t \to P} \pi_c \to \infty \tag{3}$$

$$\pi_c < 100; \qquad -\lim_{t \to E} \pi_c \to 0 \tag{4}$$

As π_c is a coefficient resulting from the division of two positive numbers, Pi-C cannot be equal to any value below zero. However, a Pi-C equal to zero translates into the worst scenario of the demise of the industry in view of the selected variables. Development of a Pi-C involves two major phases based on the proposed framework. The scope of the first phase, which is presented in this paper, was limited to development of the three dimensions of economic (E), stability (SR), and social (S) since more research and data collection were necessary for the two dimensions of development (D) and quality (Q). Fig. 2 depicts the current version of Pi-C as it gauges the pulse of the industry, and the next section provides more details on the applied variables at this phase (Table 2).

Data Collection

The selection of variables needs to ensure that the appropriate alternatives are used to gauge each factor (Nardo et al. 2005). The first phase of Pi-C involved application of the proposed framework using publicly-available data for the first three dimensions (Table 2). There were two major categories of variables: raw data and processed data. The applied variables in the form of raw data included the following:

Construction spending indicates the total dollar value of construction performed in each month in the United States and is published monthly by the U.S. Census Bureau of the Department of Commerce (U.S. Census Bureau 2014). This report is based on a survey that covers all the construction, maintenance, and rehabilitation work performed each month. The cost of labor and materials, architectural and engineering work, overhead, interest, and taxes are included. The profits of contractors also are included within the total value. Pi-C uses the not-seasonally-adjusted value of construction spending to represent the cyclical nature of the construction industry with seasonal fluctuations (Fig. 2). Using the seasonally adjusted value would result in integration of its associated assumption into the mode, but it can be used to observe short-term comparisons.

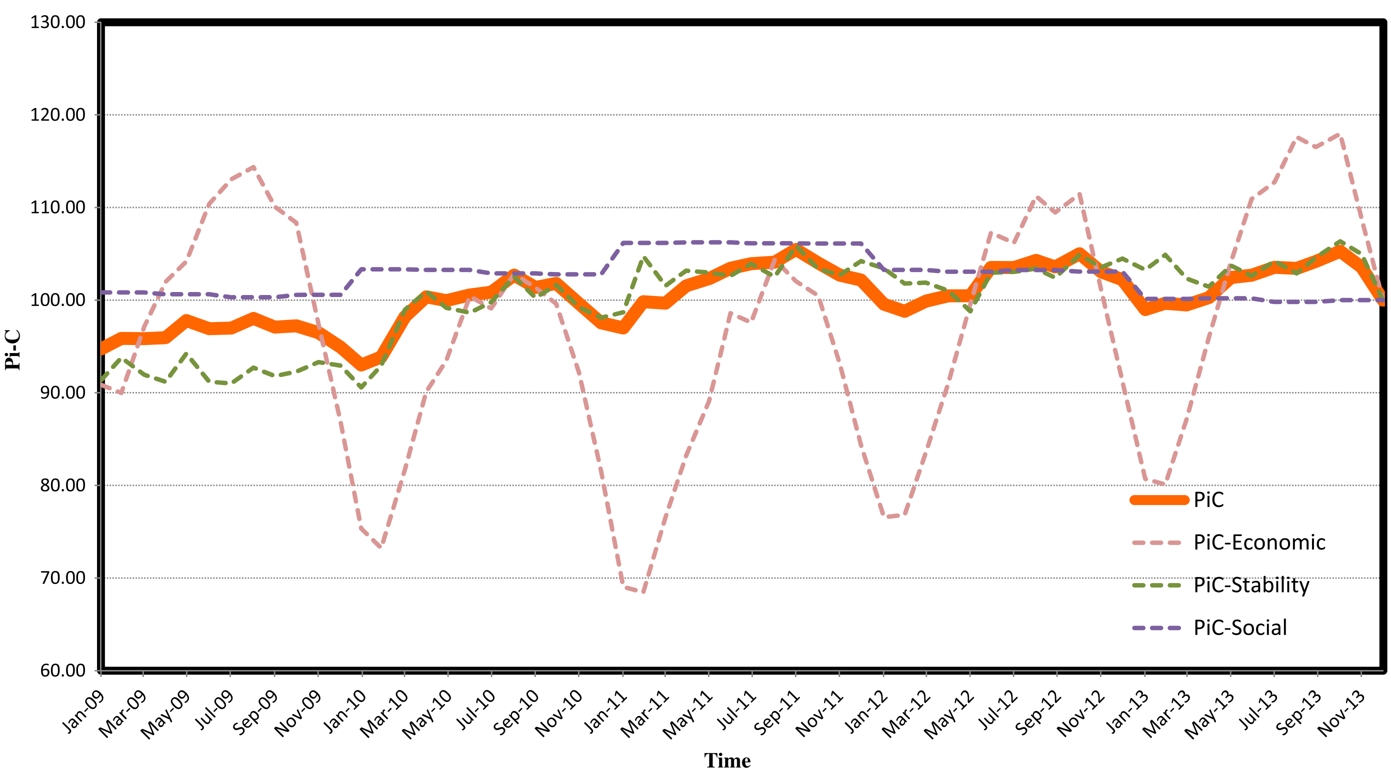

Fig. 2. Pi-C phase one including three dimensions

Table 2. Current Structure of Pi-C

Dimension	Independent indicators	Variables (I)
Economic (E)	New investment	Construction spending by DOC
Stability (SR)	Employment/layoff rate	BLS employment
	Backlog	Construction backlog indicator by ABC
	Diversity of project areas	Gini-Simpson of construction spending by DOC
	Competitiveness	Gini-Simpson of revenues from ENR Top 400 contractors
Social (S)	Safety	BLS fatal incidents data
		BLS non-fatal incidents data
	Compensation	Payscale index

Construction backlog indicator (CBI), which is published quarterly by the Associated Builders and Contractors (ABC) indicates the average months of outstanding work for the U.S. commercial, institutional, industrial, and infrastructure construction industries based on a survey of the industry (ABC 2014).

Employment statistics are elaborated by the U.S. Department of Labor Bureau of Labor and Statistics (BLS) through its Job Openings and Labor Turnover Survey (JOLTS). Within the economy, construction is the major source of investment-related employment (Barello 2014). Included in the BLS data are the total number of hires and separations for different industries. Hires include all additions to the payroll during the month such as newly hired and rehired employees; permanent, short-term, and seasonal employees; part-time employees; and transfers from other locations. It excludes transfers or promotions within the sampled establishment or subcontractors. Separations are employees separated from the payroll during the calendar month and include layoffs, quits, terminations of seasonal employees, retirements, transfers, or deaths.

In order to have a more-comprehensive view of the employment in the industry, the ratio of employment (hires) to separations were calculated for the construction industry (Fig. 3). For example, during 2009, the rate of hires did not change substantially in the construction industry while separations decreased considerably, resulting in improvement in overall employment. This ratio also was compared to the national average of the same ratio (Fig. 3). The results of the comparison indicated a net employment trend for the industry compared to the net national trend. For example, the construction industry not only improved during 2014 in terms of employment, it also outpaced the improving national employment trends (Fig. 3).

Safety data are also published annually by the BLS, which include injuries, illnesses, and fatalities for each industry sector. These statistics are based on the BLS Annual Survey of Occupational Injuries and Illnesses (SOII) (BLS 2014b) of 44 participating states including the District of Columbia as well as data from the Occupational Safety and Health Administration (OSHA) logs of workplace injuries, illnesses, and fatalities. In order to observe the trends, the ratio of the national rate to the construction rates was calculated for nonfatal and fatal incidents. The construction rate is the denominator in the ratio since higher incident rates are not desirable. This change aligned the direction of this variable with other variables.

Salary and compensation level, released as the Payscale Index (Payscale 2014), provides the earnings of full-time, private industry employees in the United States based on their total cash

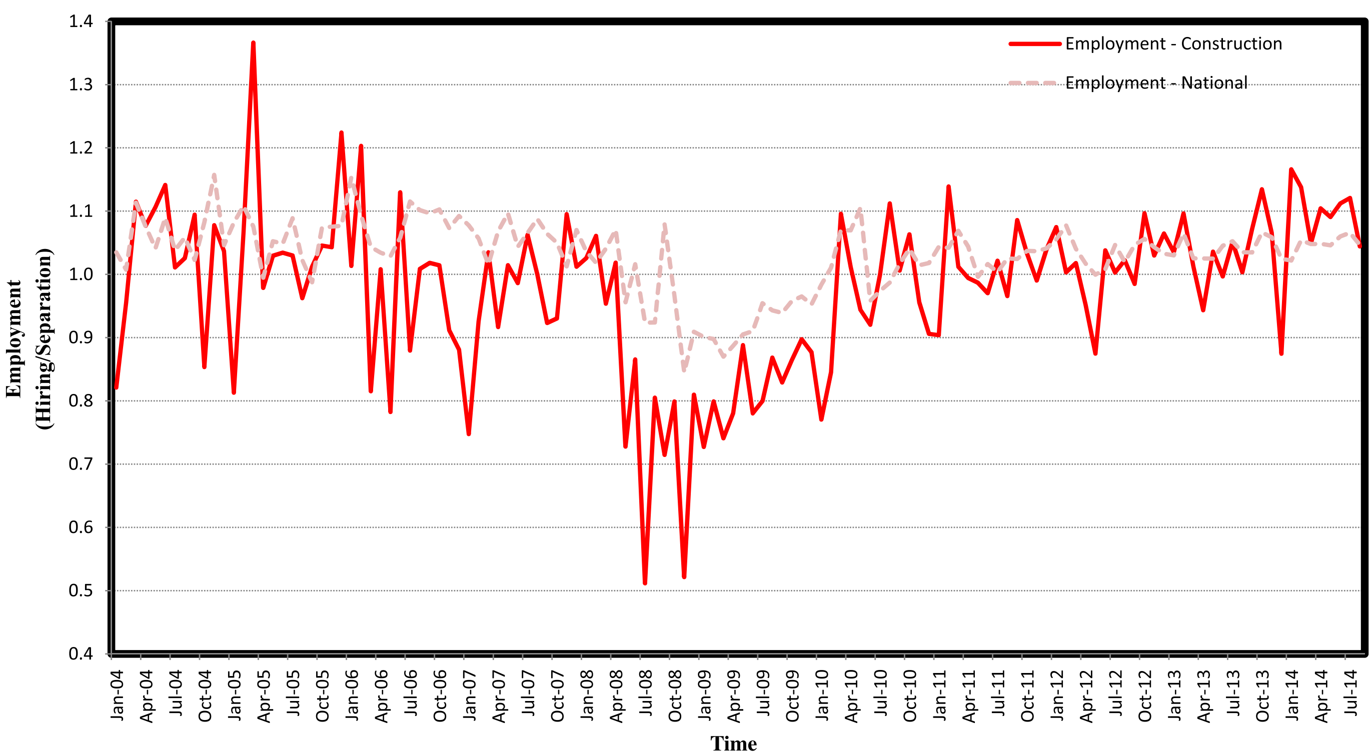

Fig. 3. Employment trend (hiring/separations) in the construction industry compared to national trend

compensation. The index assumes a 2006 average total cash compensation equal to 100 as its baseline, but it does not consider inflation. It reports earnings trends for different industries and different jobs. In order to have a comparative view of the industry, the ratio of the Payscale Index for the construction industry to the national was considered.

There were no available variables to represent some of the enumerated health factors; therefore, in addition to the previously-discussed variables, the following variables were developed for the composite index.

Competitiveness in this research aims to indicate the status of market oligopoly (Matsumoto et al. 2012; Chinowsky and Hoffman 2015) and was referred to by the surveyed industry experts as (1) competitiveness in bidding, (2) market share, and (3) availability of jobs with different sizes to allow different size companies to bid with acceptable profit margins. Therefore, competitiveness was approximated by the market share and concentration and was gauged through the distribution of revenues among the companies. The list of top 400 contractors in the United States released every year by Engineering News Record (ENR) (2014) was used to observe the distribution of companies in terms of market share (based on annual revenue). The diversity of revenues was gauged through the Gini-Simpson Index of Total Revenue of top 400 contractors, which was also suggested by Hirschman (1945) and Herfindahl (1950) as the Hirschman-Herfindahl Index (HHI) and was used at the firm level by Pitts and Hopkins (1982) and at the industry level for international construction by Zilke and Taylor (2015).

Diversity of the project areas was gauged by the breakdown of total construction spending published monthly by the U.S. Census Bureau of the Department of Commerce (Fig. 4). An overview of the breakdown of construction spending includes 16 categories of construction project areas [Fig. 4(c)]. Residential construction comprises a major proportion of the dollar value put in place each month as can be seen in the middle graph in Fig. 4(b). Diversity was gauged through the Gini-Simpson of the total construction [Fig. 4(a)], the percentage of the residential construction in the total construction, and the minimum number of specialties (project areas) that constitute more than 75% of total construction spending (Table 3).

In order to perform additional analysis on the impact of the diversity of the projects, the Gini–Simpson first was regressed against the Standard & Poor's (S&P) Building and Construction Index. The S&P Select Industry Index is an average of the measured performance of a minimum of 35 stock data from the building and construction index (S&P 2014). The results indicated different patterns of correlation for diversity rates below 0.8 and above 0.8 (Fig. 5). As the diversity rates below 0.8 were associated with the pre-2007 market crisis (Fig. 4), the negative trend observed in the diversity range above 0.8 may be attributed to the pre-existing condition of the markets and the efforts to revive it through diverse project investments. In order to confirm this breakdown, two tests were performed: the Chow test and the likelihood ratio test. In the Chow test, the hypothesis of parameter stability (i.e., estimated parameters of the regression model with the whole data do not vary with the sample size) was rejected since the F value was equal to 22.66, which was greater than the F value, 2.46, with a significance of 5%. In the likelihood ratio test, the null hypothesis that the estimated parameters are transferable regardless of the time was rejected at a 99% confidence level. The outputs of both tests confirmed the structural breakdown; therefore, a multivariate regression analysis was conducted separately for after and before 2007 using the aforementioned three different diversity metrics (Table 3) as independent variables to estimate the average S&P Building and Construction Index of each month as the dependent variable (Table 3). This model investigated the hypothesis that the diversity of project areas contributes to the future performance of the industry. The diversity of project areas is important for construction companies along with the size of the total construction spending because it suggests that companies can diversify in the case of external and internal shocks in any specific project area, such as the housing crisis in 2007, and can spread the risks (Naderpajouh et al. 2012).

As indicated in Table 4, the analysis showed a high correlation between the diversity measures and the near-future performance of the construction industry represented through one-month lagged S&P Building and Construction Select Industry Index. In addition, the Gini-Simpson of diversity was statistically significant for both analyses as it exhibited the highest *t-stat*, while the other two variables were only significant after 2007 and therefore were removed from the first analysis. The increase in the diversity of project areas may be associated with a shrinking residential market share as well as government actions to address the market crisis.

Validation of the Proposed Index

In order to ensure the statistical balance of the proposed index and to verify the interrelation of the subindices, factor analysis (FA) was conducted (Lawley and Maxwell 1971; Gorsuch 1983; Nardo et al. 2005). Since safety metrics are associated with different dynamics and the data points were very limited for this factor (only six data points for each year), they were excluded from the FA analysis. Furthermore, some data, such as the ENR top 400 contractors, are published annually while others, such as the construction backlog indicator, are quarterly. In order to perform FA, the missing data between years and quarters were reproduced assuming a linear function. However, Pi-C itself uses the same value of the variable for the whole period (year or quarter) so the effect of these variables already appears in Pi-C in a stepwise fashion. The case-to-variable ratio of FA analysis was 60:6, which satisfied the significance rule of more than 51 data-points (Lawly and Maxwell 1971). The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was equal to 0.625, and the highest variance inflation factor was equal to 4.34, less than the threshold of 5 indicated by Nardo et al. (2005).

Two latent components were suggested based on the high and moderate loadings (>0.5), while all the variables were accounted for by the first component, except the construction spending (Table 5). The weights of the subindices also were determined based on the FA as a suggested data-driven weighting. Squared factor loadings, as shown on the right side of Table 6, were applied and suggested weights were calculated based on the contribution of the variables to the overall variance (Nicoletti et al. 2000). The weights derived from FA were solely data-driven with little conceptual value (Nardo et al. 2005). Since the weights associated with each variable did not vary a great deal (except for employment), the first phase of Pi-C assumed equal weights with the option for the users to enforce suggested data-driven weights. The authors decided to move a level of compensation to the social dimension as an indicator for satisfaction of individuals. Table 2 presents the current structure of Pi-C including the developed dimensions, the health factors from Table 1, and the variables used to approximate the health factors.

Application, Interpretation, and Future Directions

Pi-C can be applied for data-driven strategy development in construction companies and data-driven policy-making at the industry level, as well as methodological exploration of future data collection needs. First, as a composite indicator, Pi-C can serve as a dashboard for the industry to observe multifaceted trends within the

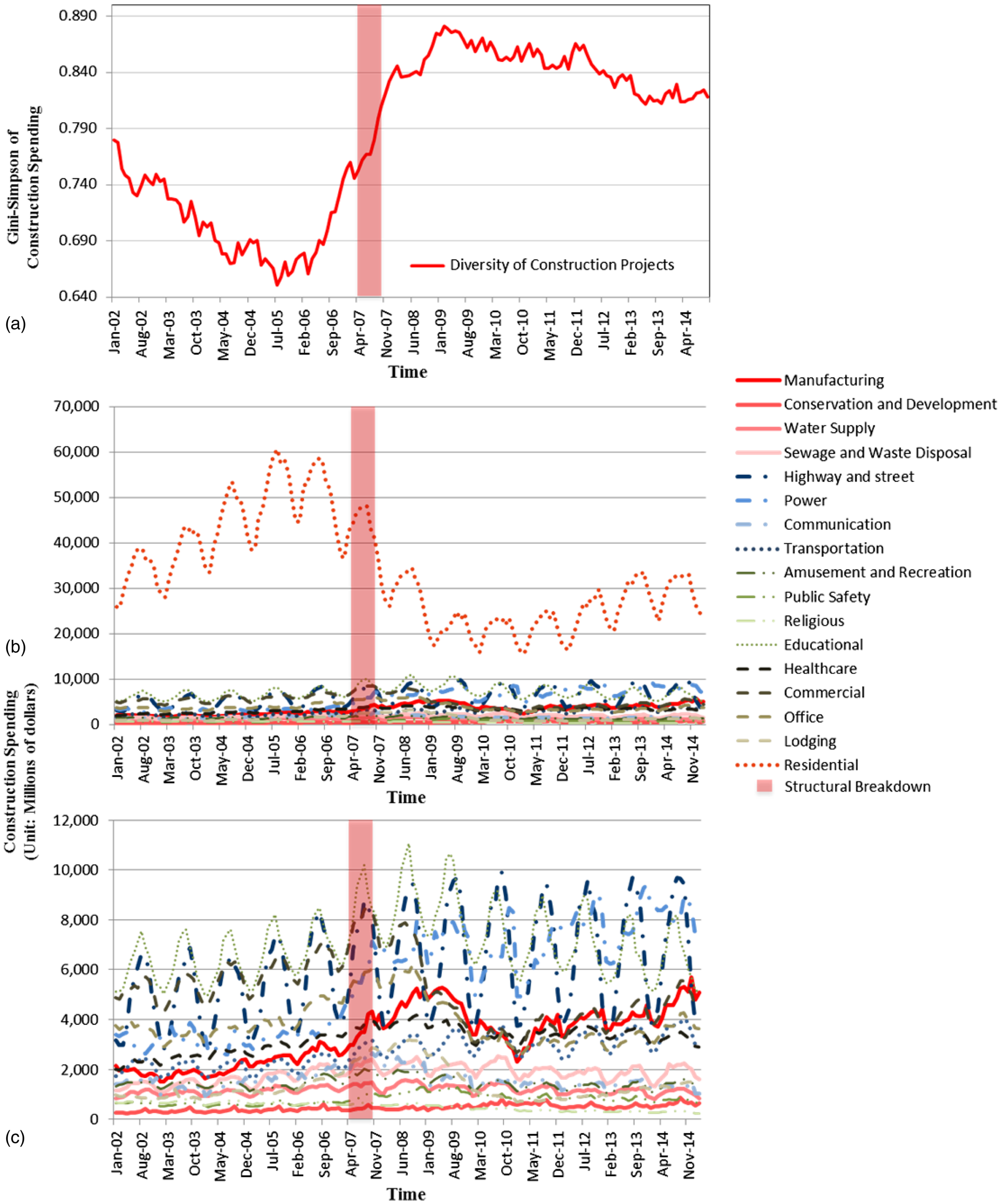

Fig. 4. Trend of diversity of construction project areas

Table 3. Summary of Independent Variables for Analyzing the Effect of Diversity

Variable description	Mean	Standard deviation	Minimum	Maximum
Gini-Simpson of construction spending	0.803	0.06991	0.6505	0.8817
Percentage of residential construction	0.388	0.09472	0.2613	0.5749
Minimum number of specialties that constitute more than 75% of total construction spending	6.605	1.1587	5	9

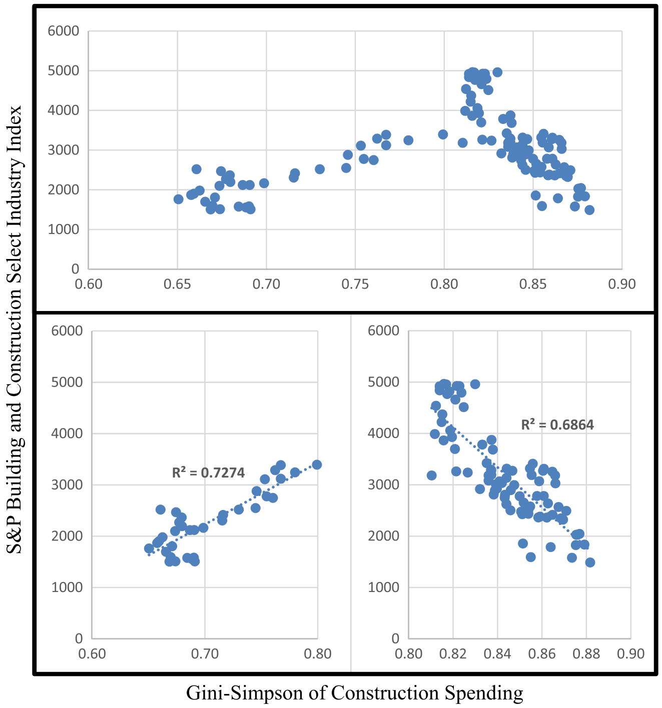

Fig. 5. Regression of diversity of project area against S&P select industry index with one-month lag

industry and decompose the index for further details. As construction is not a fully regulated market, this framework can facilitate data communication between different sectors. Periodic reports on the trends of the composite index may reflect the dynamics of the construction industry in view of the selected variables that reflect goals such as sustainability. Construction experts can apply these data in conjunction with existing sentiment indices, such as the ENR CICI, and further explore the status of the health of the industry for their strategy development for cases such as safety regulations, expansion to new project areas, and availability of jobs in each area or availability of workforce. For example, in order to cope with situations such as the 2007 market crisis, strategies could focus on changes in the focus area (Tansey et al. 2014). Pi-C, its dimensions, and variables, such as the diversity of the project areas can support development of such strategies and enable a data-driven approach to selecting the potential areas of expansion. The proposed structure also can be used with a different set of variables to gauge the dynamics of a company or a project or gauge their performance based on the appropriate set of factors.

While Pi-C facilitates decision-making at the company level, it places greater emphasis on data-driven policy-making at the construction-industry level. For example, policy-makers and regulatory bodies can use Pi-C to enforce multidimensional objectives within the industry to satisfy the requirements of sustainability and shift the focus from only the financial trends. As a result, it can be used as a data communication tool on policies for topics such as union regulations, employee benefits, diagnosing market inflations and dynamics in areas such as residential markets, or promoting new markets such as renewable energy. The Pi-C breakdown (Fig. 6) can be used to gauge the impact of policies in terms of target goals (i.e., dimensions of health). The dimensions in the example of Fig. 6 show that, although policies have not significantly improved the economic dimension, the stability dimension of health shows signs of recovery and improvement. As a result, policy-makers, such as budget planners, can further decide on channeling financial resources in different sectors of the construction, such as renewable energy, while observing its social impacts. Further exploration can lead to the source of this improvement in terms of the health factors and associated variables.

The identified health factors (Table 1) also provide a foundation to spot data-collection needs in the construction industry considering the needs delineated through the survey and availability of data. The identified gap in the availability of data in the construction industry is further illuminated by the application of Pi-C in an analysis of trends and dynamics of the construction industry as well as the importance of analytics to cope with increasing challenges at the industry level. The need for data collection in order to gauge the trajectory of the industry in terms of reference objectives, defined as health dimensions, is therefore emphasized by Pi-C. Some of these needs are being addressed in the next phase of Pi-C, including the following: refining existing dimensions with broader and richer

Table 4. Results of the Multivariate Regression Analysis for Diversity of Construction Project Areas

Time frame	R-square	Variable description	Coefficients	t-ratio	P-value
Before 2007	0.7274	Gini-Simpson of the construction spending	12,007.486	9.384	0.0000
		Constant	−6,175.344	−6.84	0.0000
After 2007	0.7217	Gini-Simpson of the construction spending	−124,410.181	−4.861	0.0000
		Percentage of the residential construction	−41,009.002	−3.205	0.0019
		Minimum number of specialties that constitute more than 75% of the total construction spending	309.147	2.546	0.0128
		Constant	119,620.468	4.722	0.0000

Table 5. Correlation Matrix for Variables Used in First Phase of Pi-C (Except Safety)

Principal component	Construction spending	Competitiveness	Backlog	Employment	Diversity of project areas	Level of compensation
Construction spending	1	−0.134	0.056	0.107	0.348	−0.077
Competitiveness	—	1	−0.861	−0.503	0.892	0.815
Backlog	—	—	1	0.523	−0.82	−0.883
Employment	—	—	—	1	−0.458	−0.566
Diversity of project areas	—	—	—	—	1	0.718
Level of compensation	—	—	—	—	—	1

Note: $N = 60$, and $p < 0.05$.

Table 6. Results of the PCA

Principal component	Prior to squared factor loadings			Squared factor loadings (scaled to unity sum)		
	Factor 1	Factor 2	Communality	Factor 1	Factor 2	PCA Weights
Construction spending	−0.054	0.989	0.912170212	0.00	0.98	0.21
Competitiveness	0.933	−0.112	0.228080932	0.87	0.01	0.18
Backlog	−0.951	−0.002	0.250145967	0.90	0.00	0.19
Employment	−0.666	0.029	0.118777129	0.44	0.00	0.09
Diversity of project areas	0.863	−0.354	0.252849483	0.74	0.13	0.15
Level of compensation	0.924	0.009	0.237976276	0.85	0.00	0.18
Explained variance	3.8991	1.036	—	—	—	—
Cumulative (%)	64.99	82.25	—	—	—	—

Note: Rotation converged in three iterations.

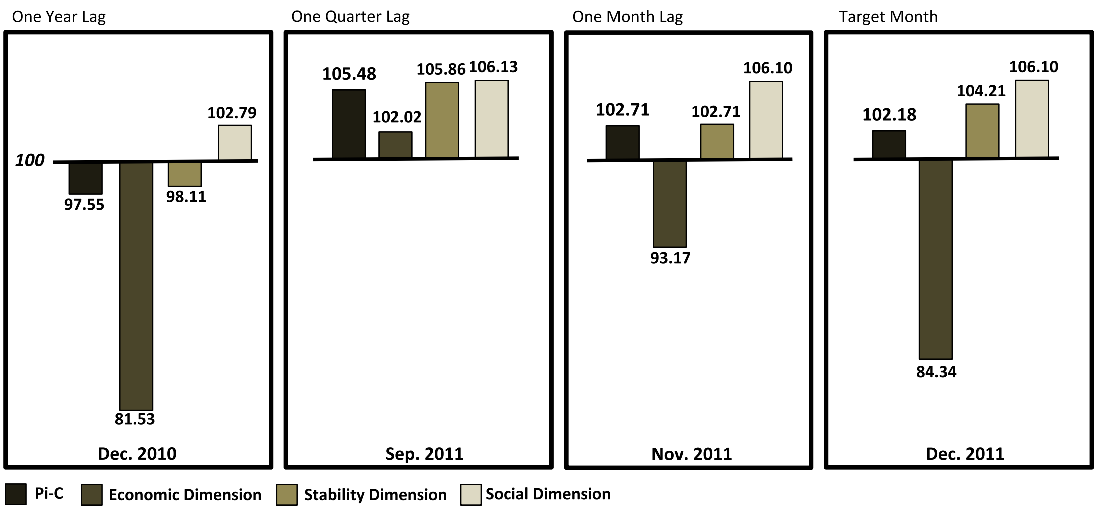

Fig. 6. Pi-C reporting example

set of data; potential revisions; development of the two remaining dimensions [i.e., development (D) and quality (Q)], as well as exploring variables that can replace the current variables to improve the quality of Pi-C. A generic procedure for future research directions includes (1) selection of the health factors based on the results of the survey on the definition of health in the industry (Table 1); (2) formulation of a measurable definition of individual health factors; (3) development of the structure of the index for specific individual health factors and associated variables; (4) validation and statistical testing for the index development; and (5) integration of the new health factors into Pi-C framework. For example, as can be seen in Table 1, there is a need for indices to reflect capacity building in terms of human resources in the construction industry. This need can be addressed through the development dimension (D) of Pi-C and the collection of data on average training rates and education levels within construction projects. Further work may also include integration of inflation through the CPI, as well as historical validation of Pi-C based on major events within the industry. Cluster analysis also can be applied to group the periods of time with similar dynamics, such as different recessions, and chronological comparison of the underlying dimensions of health in different periods. The practical validation of the indices is a dynamic process throughout the lifetime of the index. Many deficiencies of the index, as a proxy to the real-world dynamic, may be refined throughout time as a dynamic process so that the index would become a closer reflection of real-world dynamics and would better satisfy the needs of the market (Gasparatos et al. 2008).

Limitations

Regardless of the logic behind the framework of the composite index, usually the selection of components is ad hoc (McGillivray and Noorbakhsh 2004; Nardo et al. 2005; Gasparatos et al. 2008). This research faced this challenge from two fronts: (1) there is no framework to reflect dynamics of the construction industry; and (2) the phenomenon of health is not defined clearly at the industry level. This limitation of the composite indicators is justified by the research objective as it aims to promote data-driven policies and not to establish a scientific theory of causes and effects. As stated by McGillivray and Noorbakhsh (2004), when the goal of a composite index is to support policy or an advocacy-oriented approach to highlight broader issues, then the statistical redundancy of using variables that capture similar trends is less relevant. For example, in the case of Pi-C, the variable for competitiveness did not provide a significant variation. However, it was still included to reflect potential changes in different periods given the importance of competitiveness. Similarly, inclusion and exclusion of any variable should be further justified. Inclusion of health factors and their associated variables is justified in this research through the needs identified by the survey and the availability of data. Justifications for exclusion of health factors may be challenging since the number of excluded factors are infinite. In order to address this issue, the boundary of excluded factors was fixed to the survey results and the factors suggested by the industry experts. Finally, it should be noted that although the S&P Building and Construction Select Industry

Index is the only publicly-available data for the performance of construction companies, it has a limited pool from which to represent the industry as a whole. Therefore, further statistical analysis to explore the impact of diversity could be performed through more-extensive financial data from construction companies. This limitation also extends to the frequency of data as some variables, such as safety indices, are not available for each month.

Conclusions

This research intended to open the discourse on the observation and analysis of trends in the health of the construction industry. The need for the application of analytics is highlighted by the complexity of the industry, its importance in the global economy, observation of recent fluctuations, and its potential impact on the industry, as well as the increasing requirement for sustainable development within the industry. After a survey to identify the individual factors that define the health of the construction industry, the Pi-C framework was proposed as a composite index to gauge the health of the construction industry. Furthermore, the framework was tested with the publicly-available data to explore its application. Pi-C aims to address the need for observing broader issues that cover the requirements of sustainability within the construction industry. It focuses beyond the current financial status of the industry and extends its emphasis to long-term trends, the fundamental resources needed for future stability, and the capabilities required to sustain growth and development.

This effort is a continuing initiative at Purdue University and aims to provide a gateway to data-driven policy-making and strategy development in the construction industry. While Pi-C can be an indicator for the overall health of the industry, practitioners can use it to further explore the trajectories of different dimensions of health as well as its constituent health factors and associated variables. The proposed framework can facilitate the application of analytics in the decision-making process in the industry and can highlight the areas that need further data collection. The proposed index can be applied at different levels of governance to facilitate communication of trends. Researchers can also use Pi-C to explore the dynamics of this coupled system and to promote data-driven policy-making.

Acknowledgments

The authors would like to thank Professor Fred Mannering for his valuable insights and comments.

References

ABC (Associated Builders and Contractors). (2014). "Construction backlog indicator, definition and methodology." ⟨http://www.abc.org/Portals/1/Documents/CBI/CBIMethodology1.pdf⟩ (Oct. 24, 2014).

ASCE. (2007). "The vision for civil engineering in 2025." ⟨http://content.asce.org/files/pdf/TheVisionforCivilEngineeringin2025_ASCE.pdf⟩ (Jun. 29, 2015).

Ashuri, B., and Lu, J. (2010). "Time series analysis of ENR construction cost index." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0000231, 1227–1237.

Baccarini, D. (1996). "The concept of project complexity—A review." *Int. J. Project Manage.*, 14(4), 201–204.

Barello, S. H. (2014). "Consumer spending and U.S. employment from the 2007–2009 recession through 2022. Monthly labor review." ⟨http://www.bls.gov/opub/mlr/2014/article/consumer-spending-and-us-employment-from-the-recession-through-2022.htm⟩ (Dec. 23, 2014).

BLS (Bureau of Labor Statistics). (2013). "How to compute a firm's incidence rate for safety management." ⟨http://www.bls.gov/iif/osheval.htm⟩ (Oct. 12, 2013).

BLS (Bureau of Labor Statistics). (2014a). "Employed persons in nonagricultural industries by sex and class of worker." ⟨http://www.bls.gov/cps/tables.htm⟩ (Dec. 15, 2014).

BLS (Bureau of Labor Statistics). (2014b). "Survey of occupational injuries and illnesses (SOII)." ⟨http://www.bls.gov/respondents/iif/⟩ (Dec. 15, 2014).

Bon, R. (1992). "The future of international construction: Secular patterns of growth and decline." *Habitat Int.*, 16(3), 119–128.

Booysen, F. (2002). "An overview and evaluation of composite indices of development." *Soc. Indic. Res.*, 59(2), 115–151.

Bureau of Economic Analysis. (2014). "Gross output by industry." ⟨http://www.bea.gov/industry/gdpbyind_data.htm⟩ (Nov. 16, 2014).

Cao, M., Cheng, M., and Wu, Y. (2015). "Hybrid computational model for forecasting Taiwan construction cost index." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0000948, 04014089.

Chan, A. P. C., and Chan, A. P. L. (2004). "Key performance indicators for measuring construction success." *Benchmarking: An Int. J.*, 11(2), 203–221.

Chiang, Y. H., Tao, L., and Wong, F. K. (2015). "Causal relationship between construction activities, employment and GDP: The case of Hong Kong." *Habitat Int.*, 46, 1–12.

Chinowsky, P., and Hoffman, R. (2015). "Long-term viability of midsize engineering firms." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000328, 04014088.

Deng, F., Liu, G., and Jin, Z. (2013). "Factors formulating the competitiveness of the Chinese construction industry: Empirical investigation." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000161, 435–445.

De Smet, A., Loch, M., and Schaninger, W. (2007). "The link between profits and organizational performance." *McKinsey Q.*, 3, 6–9.

Dubois, A., and Gadde, L. E. (2002). "The construction industry as a loosely coupled system: Implications for productivity and innovation." *Constr. Manage. Econ.*, 20(7), 621–631.

Dutta, S. (2012). "The global innovation index 2012." *Stronger innovation linkages for global*, INSEAD and the World Intellectual Property Organization (WIPO), Fontainebleau, France.

ENR (Engineering News Record). (2014). "ENR top 400 contractors." ⟨http://enr.construction.com/toplists/Top-Contractors/001-100.asp⟩ (Dec. 12, 2014).

Esty, D. C., Levy, M., Srebotnjak, T., and de Sherbinin, A. (2005). *2005 environmental sustainability index: Benchmarking national environmental stewardship*, Yale Center for Environmental Law and Policy, New Haven, CT.

Esty, D. C., Levy, M. A., Kim, C. H., de Sherbinin, A., Srebotnjak, T., and Mara, V. (2008). *Environmental performance index*, Yale Center for Environmental Law and Policy, New Haven, CT, 382.

Fernández-Solís, J. L. (2008). "The systemic nature of the construction industry." *Architectural Engineering and Design Management*, 4(1), 31–46.

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation*, Sage Publications, Thousand Oaks, CA.

Gasparatos, A., El-Haram, M., and Horner, M. (2008). "A critical review of reductionist approaches for assessing the progress towards sustainability." *Environ. Impact Assess. Rev.*, 28(4), 286–311.

Giang, D. T., and Pheng, S. L. (2011). "Role of construction in economic development: Review of key concepts in the past 40 years." *Habitat Int.*, 35(1), 118–125.

Gorsuch, R. L. (1983). *Factor analysis*, Lawrence Erlbaum, Hillsdale, NJ.

Gregori, T., and Pietroforte, R. (2015). "An input-output analysis of the construction sector in emerging markets." *Constr. Manage. Econ.*, 33(2), 134–145.

Halpern, B. S., et al. (2012). "An index to assess the health and benefits of the global ocean." *Nature*, 488, 615–620.

Hawawini, G., Subramanian, V., and Verdin, P. (2005). "Is performance driven by industry- or firm-specific factors? A response to McNamara, Aime, and Vaaler." *Strategic Manage. J.*, 26(11), 1083–1086.

Herfindahl, O. C. (1950). "Concentration in the U.S. steel industry." Ph.D. dissertation, Columbia Univ., New York.

Hirschman, A. O. (1945). *National power and the structure of foreign trade*, University of California Press, Berkeley, CA.
Hwang, S. (2009). "Dynamic regression models for prediction of construction costs." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0000006, 360–367.
Jung, W., Han, S., Koo, B., and Jang, W. (2012a). "Which strategies are more effective for international contractors during boom and recession periods?" *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000087, 281–290.
Jung, W., Han, S., and Lee, K. (2012b). "Country portfolio solutions for global market uncertainties." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000099, 372–381.
Lawley, D. N., and Maxwell, A. E. (1971). *Factor analysis as a statistical method*, Butterworth and Co, London.
Levitt, R. (2007). "CEM research for the next 50 years: Maximizing economic, environmental, and societal value of the built environment." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2007)133:9(619), 619–628.
Manyika, J., et al. (2011). "Big data: The next frontier for innovation, competition, and productivity." McKinsey Global Institute, ⟨http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation⟩ (Feb. 5, 2015).
Matsumoto, A., Merlone, U., and Szidarovszky, F. (2012). "Some notes on applying the Herfindahl-Hirschman index." *Appl. Econ. Lett.*, 19(2), 181–184.
McGahan, A., and Porter, M. E. (1997). "How much does industry matter, really?" *Strategic Manage. J.*, 18(S1), 15–30.
McGillivray, M., and Noorbakhsh, F. (2004). "Composite indices of human well-being: Past, present, and future." World Institute for Development Economics Research, United Nations Univ., Helsinki, Finland.
McNamara, G., Aime, F., and Vaaler, P. M. (2005). "Is performance driven by industry- or firm–specific factors? A response to Hawawini, Subramanian, and Verdin." *Strategic Manage. J.*, 26(11), 1075–1081.
Momaya, K., and Selby, K. (2009). "International competitiveness of the Canadian construction industry: A comparison with Japan and the United States." *Can. J. Civ. Eng.*, 25(4), 640–652.
Naderpajouh, N., Boppana, A., and Hastak, M. (2012). "The need for indices to assess the health of the construction industry." *Construction Research Congress*, ASCE, Reston, VA, 1510–1519.
Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E. (2005). *Handbook on constructing composite indicators: Methodology and user guide (no. 2005/3)*, OECD Publishing, Paris.
Nicoletti, G., Scarpetta, S., and Boylaud, O. (2000). "Summary indicators of product market regulation with an extension to employment protection legislation." ⟨http://www.oecd.org/eco/eco⟩ (Nov. 2, 2014).
Noorbakhsh, F. (1998). "The human development index: Some technical issues and alternative indices." *J. Int. Dev.*, 10(5), 589–605.
OECD (Organization for Economic Cooperation and Development). (2013). "OECD guidelines on measuring subjective well-being." ⟨http://dx.doi.org/10.1787/9789264191655-en⟩ (Aug. 4, 2014).
Osberg, L., and Sharpe, A. (2002). "An index of economic well-being for selected OECD countries." *Rev. Income Wealth*, 48(3), 291–316.
Payscale. (2014). "The payscale index." ⟨http://www.payscale.com/payscale-index/⟩ (Nov. 14, 2014).
Phua, F. (2006). "Predicting construction firm performance: An empirical assessment of the differential impact between industry- and firm-specific factors." *Constr. Manage. Econ.*, 24(3), 309–320.
Pitts, R. A., and Hopkins, H. D. (1982). "Firm diversity: Conceptualization and measurement." *Acad. Manage. Rev.*, 7(4), 620–629.
Pries, F., and Janszen, F. (1995). "Innovation in the construction industry: The dominant role of the environment." *Constr. Manage. Econ.*, 13(1), 43–51.
Raftery, J., Pasadilla, B., Chiang, Y. H., Hui, E. C., and Tang, B. S. (1998). "Globalization and construction industry development: Implications of recent developments in the construction sector in Asia." *Constr. Manage. Econ.*, 16(6), 729–737.
Ruddock, L., and Lopes, J. (2006). "The construction sector and economic development: The 'Bon curve'." *Constr. Manage. Econ.*, 24(7), 717–723.
Rumelt, R. P. (1991). "How much does industry matter?" *Strategic Manage. J.*, 12(3), 167–185.
Schmalensee, R. (1985). "Do markets differ much?" *Am. Econ. Rev.*, 75(3), 341–351.
Schwab, K. (2013). *The Global Competitiveness Rep. 2013–2014*, World Economic Forum, Basingstoke, U.K.
Short, J. C., Ketchen, D. J., Jr., Palmer, T. B., and Hult, G. T. M. (2007). "Firm, strategic group, and industry influences on performance." *Strategic Manage. J.*, 28(2), 147–167.
Standards and Poor's. (2014). "S&P select industry indices methodology." ⟨http://us.spindices.com/indices/equity/sp-building-and-construction-select-industry-index⟩ (Dec. 21, 2014).
Stiglitz, J. E., Sen, A., and Fitoussi, J. P. (2010). "Report by the commission on the measurement of economic performance and social progress." Commission on the Measurement of Economic Performance and Social Progress, Paris.
Tansey, P., Spillane, J. P., and Meng, X. (2014). "Linking response strategies adopted by construction firms during the 2007 economic recession to Porter's generic strategies." *Constr. Manage. Econ.*, 32(7–8), 705–724.
Turin, D. A. (1978). "Construction and development." *Habitat Int.*, 3(1), 33–45.
UNDP (United Nations Development Program). (2007). "Human development report fighting climate change: Human solidarity in a divided world." New York.
U.S. Census Bureau. (2014). "Construction spending." Dept. of Commerce, ⟨https://www.census.gov/construction/c30/c30index.html⟩ (Dec. 10, 2014).
Wells, J. (1985). "The role of construction in economic growth and development." *Habitat Int.*, 9(1), 55–70.
Xu, J. W., and Moon, S. (2011). "Stochastic forecast of construction cost index using a cointegrated vector autoregression model." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000112, 10–18.
Yee, C., and Cheah, C. (2006). "Interactions between business and financial strategies of large engineering and construction firms." *J. Manage. Eng.*, 10.1061/(ASCE)0742-597X(2006)22:3(148), 148–155.
Yoon, Y., Tamer, Z., and Hastak, M. (2015). "Protocol to enhance profitability by managing risks in construction projects." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000339, 04014090.
Yu, I., Kim, K., Jung, Y., and Chin, S. (2007). "Comparable performance measurement system for construction companies." *J. Manage. Eng.*, 10.1061/(ASCE)0742-597X(2007)23:3(131), 131–139.
Zilke, J., and Taylor, J. (2015). "Evaluating the suitability of using international market analyses to characterize the global construction industry." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000312, 04014078.