1. Introduction

The importance of software affordability in early phase software design and its direct impact on software development and maintenance costs. Software cost estimation methods are employed to quantify these costs and provide stakeholders with estimates to aid in decision making. However, inaccurate cost estimation can impact software quality and development processes. The paper highlights the need for software cost estimation methods to evolve to accommodate the currently varied software development landscape and country-specific environmental and cultural aspects that may directly affect the software development effort. The paper evaluates current cost estimation datasets and identifies their suitability for application within countries with dissimilar cultural, societal, and technical constraints. The paper analyzes the characteristics of 31 software cost estimation datasets between 1981 and 2017 and compares their attributes to demonstrate the need for explicit representation of environmental factors within future datasets.

2. Dataset

The study used 31 datasets for algorithmic cost estimation models, which were a combination of datasets from the Open Science tera-PROMISE repository, datasets surveyed by Mair et al., and datasets in the study by DE jaeger et al. Private datasets with publicly available attributes were excluded. The date and sector of the datasets were determined using various criteria, and datasets with comprehensive attributes were chosen. The characteristics of all the datasets used are detailed in Table I in chronological order.

A. Dataset Characteristics

The majority of datasets used for software cost estimation were collected during the 1980s and 1990s, and only a few recent datasets are available. 48% of the datasets are from Europe, 31% from North America and 21% from Asia and Australia. No datasets represent projects from Africa or South America, and

no low-to-middle income countries from Asia are represented. The datasets come from various sectors, but there is a lack of details on the comprehensiveness of the dataset application areas. The number of attributes within datasets is varied, with an overall average of 13 attributes.

However, the diversity of projects and attributes makes comparative costing studies difficult, and most studies report inconclusive results. Most datasets represent outdated project development environments and exclude environments with dissimilar cultural and societal practices, which is a limitation given the significance of the software industries in developing and emerging economies.

B. Cost Attributes

The text describes a study on the attributes of 30 datasets related to project cost estimation in software engineering. The authors categorized the attributes into six categories: general information, users, developers, size, project, and product. They combined attributes with similar meanings and reduced the total number of attributes to 48. The text reports the distribution of attributes in each category for all datasets and compares the mean coverage of each category between datasets collected before and after the year 2000.

The study found that product size was an important attribute included in all datasets, but only a few datasets contained attributes from the users category, which limits their applicability in low-to-middle income countries. The developers category is an important cost factor in technically-constrained environments, but the team continuity attribute is not considered in most datasets. The KSLOC and function points attributes were the most common attributes within the size category, and the programming language and tool availability attributes were the most common attributes in the project category. The product complexity and reliability attributes were the most common attributes in the product category.

C. Importance of Enviromental Factor

Accounting for country-specific factors and environmental differences is essential for accurate software cost estimation. When using generic datasets or attributes at an international level, it is important to consider environmental factors and their impact on cost estimation. A method is needed to determine their impact on other cost drivers.

3. Solution

We discuss the importance of accurate software cost estimation in the context of internationalization, outsourcing, and a growing international software market. It notes that while most studies use cost estimation datasets to build and compare different software cost estimation methods, there is limited research on the impact of cost drivers on the accuracy of cost estimation methods. It also highlights the need to consider environmental factors within cost drivers and datasets, including factors that differ from country to country, such as cultural and organizational norms. It provides examples of studies that have examined the impact of environmental factors on software cost estimation, including those related to global software development. Further research is needed to identify and impact the identification and impact of these factors.

4. Conclution

This paper analyzed publicly available software cost estimation datasets to determine their inclusion of environmental factors that impact software development costs. It found that most datasets represented US and European organizational environments, and the datasets from 2000 and later included more diverse attributes and increased coverage of organizational environmental factors. However, none of the datasets considered country-specific environmental cost factors. Further research is needed to identify and quantify the impact of environmental factors on software cost estimation, including case studies within different environments and countries.