

Duration and cost variability of construction activities: an empirical study

Article

Accepted Version

Ballesteros-Pérez, P., Sanz-Ablanedo, E., Soetanto, R., González-Cruz, M.C., Larsen, G. and Cerezo-Narváez, A. (2020) Duration and cost variability of construction activities: an empirical study. *Engineering Construction and Architectural Management*, 146 (1). ISSN 1365-232x doi: [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001739](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001739)
Available at <http://centaur.reading.ac.uk/85395/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <https://ascelibrary.org/doi/full/10.1061/%28ASCE%29CO.1943-7862.0001739>

To link to this article DOI: [http://dx.doi.org/10.1061/\(ASCE\)CO.1943-7862.0001739](http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0001739)

Publisher: Wiley Blackwell

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

1 On the duration and cost variability of construction activities: an empirical study

2 Ballesteros-Pérez, P. Ph.D.^{1*}; Sanz-Ablanedo, E. Ph.D.²; Soetanto, R. Ph.D.³;

3 González-Cruz, M^a.C. Ph.D.⁴; Larsen, G.D. Ph.D.⁵; Cerezo-Narváez, A. Ph.D.⁶

4 Abstract

5 The unique nature of construction projects can mean that construction activities often
6 suffer from duration and cost variability. As this variability is unplanned it can present a
7 problem when attempting to complete a project on time and on budget. Various factors
8 causing this variability have been identified in the literature, but they predominantly refer to
9 the nature and/or context of the whole project, rather than their specific activities.

10 In this paper, the order of magnitude of and correlation between activity duration and
11 cost variability is analyzed in 101 construction projects with over 5000 activities. To do this,
12 the first four moments (mean, standard deviation, skewness and kurtosis) of actual versus
13 planned duration and cost (log) ratios are analyzed by project, phase of execution and activity
14 type. Results suggest that, contrary to common wisdom, construction activities do not end
15 late on average. Instead, the large variability in the activity duration is the major factor
16 causing significant project delays and cost overruns. The values of average activity duration
17 and cost variability gathered in this study will also serve as a reference for construction

^{1*} Senior researcher. Dpto. de Ingeniería Mecánica y Diseño Industrial, Universidad de Cádiz.

Avda. Univ. de Cádiz 10, Puerto Real, 11519 Cádiz, Spain. Phone: +34 956 483 200, pablo.ballesteros@uca.es

² Associate Professor. Dpto. Tecnología Minera, Topográfica y Estructuras. Universidad de León.

Avda. Astorga, s/n. 24400 Ponferrada, León. Spain. Phone: +34 987 442 110, esana@unileon.es

³ Senior Lecturer. School of Architecture, Building and Civil engineering. Loughborough University

Loughborough, LE11 6DF, UK, United Kingdom. Phone: +44 (0) 1509 228 748, r.soetanto@lboro.ac.uk

⁴ Associate Professor. GIDDP, Depto. de Proyectos de Ingeniería. Universitat Politècnica de València.

Camino de Vera s/n, 46022 Valencia, Spain. Phone: +34 963 879 866, mcgonzal@dpi.upv.es

⁵ Associate Professor. School of the Built Environment. University of Reading

Whiteknights, Reading RG6 6DF, United Kingdom. Phone: +44 (0) 118 378 7185, g.d.larsen@reading.ac.uk

⁶ Assistant Professor. Dpto. de Ingeniería Mecánica y Diseño Industrial, Universidad de Cádiz.

Avda. Univ. de Cádiz 10, Puerto Real, 11519 Cádiz, Spain. Phone: +34 956 483 311, alberto.cerezo@uca.es

18 managers to improve future construction planning and project simulation studies with more
19 realistic data.

20 **Keywords:** scheduling; activity variability; merge event bias; network topology; project delays

21

22 **Introduction**

23 Construction activities usually suffer from variability in terms of both duration and
24 cost. With each construction project being unique, factors of this variability are plentiful
25 (Ballesteros-Pérez et al. 2017). These factors include project location, clients, regulations,
26 labor, equipment, technology, subcontractors, experience, stakeholders, even the project
27 team, are likely to change, at least partially, among projects (Chudley and Greeno 2016). All
28 these factors, plus many other, make of the duration and cost estimation exercise, a
29 challenging task for construction managers.

30 It may be easy to believe, though, that construction activities are apparently more
31 likely to end later and cost more than the other way around. In fact, this would constitute a
32 compelling reason why so many construction projects end late and exceed their initial budget.

33 Factors that cause projects to end late or result in cost overruns have been studied in
34 the construction literature for a long time. Some of the most recurrent are poor planning and
35 control practices, deficient construction site management, shortages of labor and/or low
36 productivity, problems with the supply chain and/or procurement practices, contractor's
37 and/or client's financial problems, project specifications or design changes, communication
38 and/or co-ordination problems among stakeholders, interferences with onsite services,
39 adverse weather conditions, and legal disputes and contract claims (Ballesteros-pérez et al.
40 2015, 2018b). Among all these, however, poor planning and control practices are consistently
41 among the most pervasive (AlSehaimi and Koskela 2008).

42 Ballesteros-Pérez et al. (2018a) recently showed how the most common scheduling
43 techniques (Gantt chart, Critical Path Method and Project Evaluation and Review Technique,
44 PERT) consistently underestimate the actual project duration and cost. One of the major
45 causes of this underestimation came precisely from neglecting activity duration variability.

46 Apart from the classical scheduling techniques, more advanced techniques for getting
47 improved project duration and/or cost estimates have been proposed over the years (e.g.
48 fuzzy logic, neural network analysis, Monte Carlo simulations, artificial intelligence methods,
49 many variants of PERT, and even more extensions of Earned Value Management
50 (Ballesteros-Pérez, 2017a)). What all these methods have in common, classical and modern
51 alike, is that they all require some prior estimates of the potential activity durations and costs.
52 For example, PERT-related techniques generally resort to three-point estimates (pessimistic,
53 optimistic and most likely durations and costs); Monte Carlo simulations require the
54 statistical distributions of each activity as input; and neural network analysis and artificial
55 intelligence methods require training sets of similar construction projects. Access to this
56 information is often the major limitation of these methods. Similarly, realistic data on the
57 correlation between activity duration and costs is also a rare commodity, which forces these
58 techniques to either assume independence between activities and costs, or resort to subjective
59 correlation factors (Banerjee and Paul 2008; Cho 2009). Consequently, when enough quantity
60 or quality of information is not available, the forecasting accuracy of the actual project
61 duration and/or cost is expected to be unreliable.

62 Unfortunately, despite its importance, there is a dearth of research into activity duration
63 and cost variability in the construction management literature. Maybe, the only exception
64 would be the work of Trietsch et al. (2012) who attempted to establish a distribution that
65 satisfactorily describes construction activity durations. However, as early suggested by
66 MacCrimmon and Ryavec (1964), trying to find a universal distribution that fits all types of

67 activities is a futile effort because each type of activity is unique. Furthermore, its context
68 might also have a significant influence which is difficult, if not impossible, to parameterize
69 mathematically.

70 Nonetheless, these difficulties should not be a deterrent to, at least, attempting to measure
71 the average level of variability of construction activity durations and costs. As argued, this
72 would be an extremely valuable input for future project duration and cost forecasting
73 techniques, as well as providing powerful baseline information for enhancing project control
74 and monitoring.

75 Hence, the present paper precisely attempts to fill this research gap in the construction
76 management literature: measure the average level of activity duration and cost variability. It
77 will also justify how and why, given this level of activity variability in common project
78 networks, it is expected that most construction projects end late and go over budget. To
79 achieve this, the actual/planned (log) ratios of many project and activity durations and costs
80 will be analyzed. The correlation between activity durations and costs will also be studied.
81 Finally, the most common network topologies (descriptors of what the project networks are
82 like, that is, how activities are arranged and connected with each other) will be summarized
83 and the potential impact of activity variability on these networks described in detail.

84 The paper will be structured as follows. The *background* section will provide an
85 overview of the importance of the first four moments of the activity duration and costs
86 impacting the final project duration and cost. This section will introduce the concept of merge
87 event bias and describe how it may cause project delays and cost overruns depending of each
88 project network topology. The *materials and methods* section will describe how a dataset of
89 101 projects was classified according to different activity categories, and then their log actual
90 vs planned durations and cost deviations analyzed activity by activity. The *discussion* section

91 will provide insights on to what the numerical results mean and how they are connected to the
92 project network topology in common construction projects. Finally, the *conclusions* will
93 summarize the whole analysis, highlight the major contributions to the body of knowledge,
94 state the study limitations, and propose future research continuations.

95

96 **Background**

97 There have been numerous studies analyzing delays and cost overruns in construction
98 projects at project level (e.g. (Hamzah et al. 2011; Keane and Caletka 2008; Mahamid et al.
99 2012; Ogunlana et al. 1996; Orangi et al. 2011; Senouci et al. 2016)). Most studies have
100 focused on either establishing the causes of delays and cost overruns, and/or proposing some
101 regression analyses to avoid slippages in the future. Generally, these studies have been
102 aligned with a more reductionist perspective, seeking to emphasize a particular context (same
103 region, client, type of projects, or a combination of these).

104 Conversely, there have not been hardly studies measuring the ‘activity’ durations and
105 costs, let alone their variability in real construction projects. With the exception of Trietsch et
106 al. (2012) mentioned earlier, perhaps the closest are a handful of studies analyzing the
107 sensitivity of the project duration to different levels of activity mean duration and dispersion
108 (e.g. Elmaghraby & Taner (1999) and Elmaghraby (2000)).

109 Additionally, but from a purely mathematical and simulation perspective, some
110 studies have tried to gauge to what extent the adopted activity statistical distributions have a
111 significant repercussion on the final project duration. In this regard, a recent study by Hajdu
112 and Bokor (2014) concluded that the maximum project duration deviation when using
113 alternative activity distributions was generally well below 10%. This finding resonated with
114 observations from an earlier study on the limitations of PERT. MacCrimmon and Ryavec

115 (1964) showed that, if triangular distributions for modelling activity durations had been
116 chosen instead of Beta distributions, the probabilistic project duration would have produced
117 almost identical results.

118 The reason why the choice of a particular statistical distribution does not seem that
119 relevant is because the third and fourth moments (skewness and kurtosis) are blurred very
120 quickly in Stochastic Network Analysis (SNA) (Hajdu and Bokor 2016). At the time of
121 writing, SNA is considered the most accurate approach to model project schedule networks
122 (Ballesteros-Pérez, 2017b). In SNA, activity durations and costs are modelled by statistical
123 distributions (with or without correlation with each other). More precisely, distributions are
124 summed when computing the total costs of activities, or the total duration of activities
125 arrayed in series. On the other hand, the maximum of distributions (instead of a sum) is
126 calculated whenever we calculate the total duration of a set of activities placed in parallel. In
127 either case, the third and fourth moments (skewness and kurtosis) have a minor influence on
128 the resulting distribution (of a path or project duration).

129 However, the first two moments (mean and variance, or alternatively, standard
130 deviation) play a major role in the resulting distribution modelling the total project duration.
131 When there is some correlation between durations and costs (virtually always in construction
132 projects), they also have an indirect but still significant, influence on the final project cost.

133 To sum up, when two or more distributions are convoluted (summed for computing
134 the project cost or the duration of activities in series) the resulting distribution, by the Central
135 Limit Theorem, quickly converges to a Normal distribution. The mean and variance of this
136 Normal distribution correspond to the sum of means and variances, respectively, of the
137 individual activity distributions. Therefore, the first two moments will mostly determine what
138 the resulting distribution looks like.

139 When some activities are arranged in parallel and they all need to finish before the
140 project can continue, the resulting distribution quickly converges to an extreme value
141 distribution of maxima (normally a Fréchet or a Gumbel distribution) (Dodin and Sirvanci
142 1990). Again, the first two moments of the involved activity distributions will determine the
143 location and scale of the resulting extreme value distribution. This phenomenon is commonly
144 known as the ‘merge event bias’ (Khamooshi & Cioffi, 2013; Vanhoucke, 2012) and it is
145 indeed the major source of inaccuracy of all deterministic scheduling techniques.

146 Real construction project schedules (networks) generally involve many subsets of
147 activities both arranged in parallel and in series. Hence, multiple convolutions (sums) and
148 maxima of distributions need to be computed so that the final project duration and cost can be
149 calculated. The influence of each activity’s first two moments (mean and standard deviation)
150 will be key in this final result. This justifies why an order of magnitude of these two moments
151 is worth collecting from a representative dataset of real construction activities.

152 Finally, another factor that determines how the activity distributions are merged with
153 each other is dependent on the project network topology itself. Network topology refers to the
154 logical layout of a network (a project schedule). It defines the way different activities (often
155 referred to as nodes) are placed and interconnected with each other. Many metrics have been
156 proposed for describing the network configuration. Some well-known examples are the
157 Coefficient of Network Complexity (Davies 1973; Pascoe 1966), the Order Strength (Mastor
158 1970) and the Complexity Index (Bein et al. 1992). However, these only capture the project
159 complexity and will not be used here.

160 Instead, this study will make use of four topology measures that describe the structure
161 of an activity-on-the-node network, not just its complexity. These measures were initially
162 proposed by Tavares et al. (1999) and later improved by Vanhoucke (2008). The four
163 measures (also named *indicators*) used are: serial-Parallel (SP) indicator, Activity

164 Distribution (AD), Length of Arcs (LA) indicator, and Topological Float (TF) which will be
165 explained in the following sections. All these indicators range between 0 and 1 and constitute
166 simple measures describing to what extent the first two moments of the construction activities
167 may condition the final duration and cost of a project.

168

169 **Materials and methods**

170 In this section, the characteristics of the projects and activity datasets analyzed are
171 described first. The details of how the activity and project data was filtered and categorized,
172 under multiple levels of analysis, is also presented. Next, the first four moments of activity
173 durations and costs are reported and commented separately. Finally, the correlations between
174 activity durations and costs are reported along with their statistical significance.

175

176 ***Projects and activities dataset***

177 This research used two different project datasets. The first (and main) one is analyzed
178 at both activity- and project-level. The second dataset contains project level information
179 (planned and actual project durations and costs) and will be used for illustrative purposes in
180 the *discussions*.

181 In order to obtain representative values of the first four moments of the activity
182 durations and costs, a significant amount of activities is necessary. In the first dataset, 101
183 construction projects are analyzed initially encompassing 5,697 activities.

184 Projects are classified in four types: Building, Civil engineering, Industrial and
185 Services. Building projects are mostly aimed at constructing a building or parts of a building.
186 Civil engineering refers to infrastructure construction in general. Industrial projects refer to

187 installations and/or electromechanical equipment. Services refer to projects with a significant
188 operational and/or production component.

189 The 101-project dataset was retrieved from a real projects dataset originally
190 developed by Batselier and Vanhoucke (2015) and Vanhoucke et al. (2016). Although the
191 exact location of those projects is not disclosed in most cases (due to a confidentiality clause
192 with the information donors), it is known that most of them belong to Belgium, the
193 Netherlands, Italy, USA and Azerbaijan.

194 At the time of writing, the complete project dataset is curated by the Operations
195 Research & Scheduling Research Group at Ghent University and comprises 125 projects. 24
196 projects out of the 125 were not used as they did not include tracking information (actual
197 activity durations and costs). All 125 projects, however, can be accessed at the website of
198 OR-AS.be (2018). The major features of the 101 construction projects selected for this study
199 are summarized in Table 1. The last four columns of Table 1 include some project network
200 topological information (indicators SP, AD, LA, and TF) that will be used later.

201 **<Insert Table 1 here>**

202 We deem the variety and number of project types, costs, durations, topologies and
203 number of activities as sufficiently representative for a first representative analysis. Yet,
204 further details and specific project information can also be found as individual project cards
205 at OR-AS.be (2018).

206

207 *Analysis outline*

208 This analysis focuses first on the activity-level deviations of durations and costs.
209 Project-level data will also be analyzed later, but from a complementary point of view to

210 activities analyses. The activity duration and cost deviations are calculated for each activity i
211 in the first dataset according to these two expressions, respectively:

$$212 \quad \textit{Activity duration deviation of activity } i = \text{LOG}_{10} \left(\frac{\textit{Actual duration of activity } i}{\textit{Planned duration of activity } i} \right) \quad (1)$$

$$213 \quad \textit{Activity cost deviation of activity } i = \text{LOG}_{10} \left(\frac{\textit{Actual cost of activity } i}{\textit{Planned cost of activity } i} \right) \quad (2)$$

214 It is worth emphasizing that both ratios above are expressed in logarithmic scale. This
215 is important, as ratios of variables which are always positive (e.g. durations and costs) are not
216 symmetrical respect to the value 1. The scale distortion of these ratios (they range between 0
217 and 1 when the denominator is bigger than the numerator, but between 1 and + infinity when
218 the numerator is bigger than the denominator) creates an artificial positive skewness in the
219 data distribution that can only be removed by taking the log ratios beforehand. Additionally,
220 in log scale, the variable variances are additive, rather than multiplicative.

221 Therefore, we will take the logarithm of every ratio before analyzing their activity
222 duration and cost moments. We resorted to logarithms with base 10 because their orders of
223 magnitude are a little more familiar, but any other base would have been possible.

224 Lastly, it is important to note that ratios in natural scale from 0 to 1 correspond to
225 values from -infinity to 0 in any log scale. Whereas ratios in natural scale from 1 to +infinity
226 correspond to the $(0, +\infty)$ range. Both ranges also have a symmetrical correspondence with
227 each other in log scale (e.g. ratios $\frac{1}{2}$ and 2 in natural scale have the same values with opposite
228 signs in log scale, that is -0.301 and 0.301, respectively) which makes the interpretations of
229 variability results easier. Bearing this in mind, the next step consists of describing how the
230 activities were grouped to analyze their ratios and produce robust results. The progressive
231 classification levels can be found in Table 2.

232

<Insert Table 2 here>

233

From top to bottom, three levels of activity classifications are presented. Each level

234

consists of three types of activities:

235

- Planned and Performed (P&P). These activities correspond to activities that were

236

initially planned and were also finally executed in the projects analyzed. These are the

237

most frequent and the only ones that are considered in the analysis.

238

- Unplanned but Performed (UbP). These activities correspond to activities that were

239

not initially planned but that were deemed necessary and had to be eventually carried

240

out. These activities were removed from the analysis because their ratios converged to

241

+ infinity (as the planned values in the denominators equal 0), and because most of

242

the time they come from planning mistakes or omissions.

243

- Planned but not Performed (PbnP). These activities correspond to activities that were

244

initially planned, but that were not executed in the end. These activity ratios would

245

equal zero in natural scale but their logarithmic values would converge to - infinity.

246

They also represent bad estimates of the planned schedule like UbP activities, hence,

247

they were also removed from the analysis.

248

Concerning activity grouping, four levels of analysis (0 to 3) were considered:

249

- Level 0 comprises all activities analyzed from all projects. This allows drawing

250

general average conclusions without paying attention to proportions nor types of those

251

activities.

252

- Level 1. Activities are classified under the same four types of projects stated in Table

253

1 (building, civil engineering, industrial and services). As expected, this level allows

254

analyzing how the activity durations and costs deviations differ by (generic) types of

255 projects. Some group average and dispersion results of activity durations and costs are
256 also included for reference on the right columns of Level 1 sub-table.

257 • Level 2. Within the previous four project type categories we further classify activities
258 into three standard phases of the every project lifecycle according to the PMBoK:
259 Planning, Execution and Closure (Project Management Institute 2017). Classifying
260 activities into these three categories is straightforward with the activity descriptions
261 available in almost all projects. The fourth phase considered by the PMBoK
262 (Monitoring and control) is not relevant for this analysis, therefore not considered.

263 • Level 3. For the *execution* phase of *Building* and *Civil engineering* projects only
264 activities are further classified into five generic groups, called here *activity types*
265 (auxiliary works, substructure, superstructure, specialized works, and facilities).
266 These are also common and relatively straightforward groups of activities in most
267 construction projects. For a more detailed description of the scope of each group the
268 reader is referred to Chudley and Greeno (2016).

269 Level 3 allowed classifying activities into one last level right above the nature of the
270 activity itself. Activities in this level were classified mostly thanks to the descriptions of the
271 project *summary activities* (that were indeed not used for anything else in the analysis).

272 Finally, as highlighted at the beginning of level 3, only activities from the *execution* phase of
273 *building* and *civil engineering* projects were used. This is due to the number of *execution*
274 activities in *Industrial* projects being considered too low. Also, because *Execution* activities
275 belonging to *Services* projects, despite higher in number, were found too heterogeneous. The
276 latter made hard to classify these activities within similar self-contained categories (*Services*
277 projects are indeed much more varied regarding the nature of its activities).

278

279 ***Activity duration results***

280 The first four moments (average, standard deviation, skewness and kurtosis) of the
281 activities log ratios were analyzed according to the four levels described in Table 2. Table 3
282 shows now the results for the activity *duration* log ratios ($\text{LOG}_{10}(\text{actual} / \text{planned})$).

283 **<Insert Table 3 here>**

284 For each case and level analyzed, four numerical values are displayed: n (the sample
285 size, that is, the number of activities used to calculate the four moments), and the four
286 moment values (in logarithmic scale). However, due to the major relevance of the first two
287 moments (average and standard deviation) these two have also been included in natural scale
288 within parentheses right below their respective logarithmic values. Values in natural scale are
289 expected to help the reader to better grasp the order of magnitude of these moments. With
290 this information, Table 3 is self-explanatory. The number of readings and details in this table
291 are numerous, so attention is given to the most relevant findings.

292 Concerning *Averages*, it is striking to observe how most values remain very close to 0
293 (in log values) or 1 (in natural values). Some exceptions may be *Services* projects and the
294 *Planning* phase activities (Level 2) from *Building* and *Civil engineering* projects. Yet, in the
295 latter, average ratios values remain close to 5% (in log values) or 11% (in natural values).
296 Overall, as these log ratios are so close to zero, this suggests that construction activities do
297 *not* end late (on average). This may be an unexpected finding, as the easier explanation for
298 projects ending late was that its activities ended late on average. This result seems to suggest
299 the problem lies somewhere else.

300 Concerning the *Standard Deviation* (SD) values, results are very different. SD, by
301 definition, can only be positive but it is quite clear that, unlike the averages, SDs are not close
302 to zero. Instead, with a few exceptions, SD values are almost always above 0.15 (in log scale)
303 or 43% (in natural scale) between the actual and planned durations. This is an extremely high

304 level of variability and, despite construction activities do not end late *on average*, they do
305 suffer from wide dispersions which condition to a big extent the project-level delays, as will
306 be justified later. On a secondary note, *Industrial* and *Services* projects also have a bigger
307 variability than the other types of projects. Interpretations by project phase (level 2) and
308 activity type (level 3) are more varied.

309 The results on *Skewness* are relatively uniform. A common rule of thumb assumes
310 that skewness values ranging from -2 to +2 are indicative of a low distribution asymmetry
311 (George and Mallery 2010). This is the case in Table 3 with very few exceptions. Therefore,
312 the log ratios distribution must be approximately symmetrical and, combined with averages
313 also close to zero, we can conclude that there is approximately the same probability of
314 finding early activities than tardy activities.

315 Concerning *kurtosis*, the picture is very different. Values are generally well above 3,
316 which would describe the kurtosis corresponding to the Normal distribution. This result
317 means that log ratio duration values resemble a peaked distribution with heavy tails. In other
318 words, the majority of the actual durations are not close to their planned values. As stated
319 earlier, many other readings may be extracted from Table 3. However, for the sake of clarity,
320 only the most relevant high-level interpretations are presented.

321

322 ***Activity cost results***

323 Table 4 represents the first four moments of the activity actual versus planned *cost* log
324 ratios. In parentheses, we can find the antilogarithmic (natural scale) values of the first two
325 moments as well. Table 4 values differ substantially from those found in Table 3.

326

<Insert Table 4 here>

327 Concerning *Average* values, most of them are clearly positive and generally above
328 1.01 (in log values) or alternatively above 3% (in natural scale). A clear exception may be the
329 *Industrial* projects whose average is negative. This may be because *Industrial* projects are
330 frequently composed of electromechanical equipment whose procurement prices are
331 relatively easier to estimate more accurately ex-ante than other types of projects.
332 Additionally, *Civil engineering* and *Services* projects are among the ones whose activities
333 tend to suffer from more cost overruns. This may be due to *civil engineering* projects being
334 (generally) less standard than Buildings whose average log ratios remain closer to 0. On the
335 other hand, *services* projects as indicated in Table 3, suffered from more delays on average
336 than other types of projects. Being these types of projects frequently more labor intensive, it
337 seems logical that those extra durations are correlated with these extra costs.

338 Concerning *Standard Deviation* (SD), variability is even more evident than in the case
339 of duration log ratios. On level 0 we can appreciate how the average activity SD reaches 0.25
340 (78% of variability in natural scale). On level 1, no project type has a variability below 0.16
341 (46% of variability, in the case of *Building* projects) and two of them (*Civil engineering* and
342 *Services*) remain above 0.30 (>100% of variability). SDs on levels 2 and 3 offer similar
343 readings but with wider values.

344 Concerning *skewness*, cost log ratios are more varied than their duration counterparts.
345 In general, when *average* values are negative, the skewness values are also predominantly
346 negative. Similarly, when the *average* costs are positive, the cost distribution is also
347 positively skewed.

348 Concerning *kurtosis*, values are much higher than its duration ratios counterpart too.
349 This would be indicative again that most activity actual costs substantially differ from their
350 planned values (a high proportion of the actual costs tend to be substantially different from
351 their planned costs).

352

353 ***Activity duration and cost correlation***

354 Numerical results of the log ratios of the first four moments offered very interesting
355 information about the nature of duration extensions and cost overruns at activity level. It is
356 not the intention of this study to find a distribution that fits these four moments, though. As
357 suggested by other researchers and also discussed earlier, each activity is different in nature
358 and it is quite likely that a fit-for-all distribution does not exist. Indeed, on observing the wide
359 range of skewness and kurtosis values in Tables 3 and 4, that seems to be exactly the case.

360 However, a pending but also equally relevant issue is to analyze the potential
361 correlation between activity duration variation and cost variation. For this aim, all activities
362 were grouped under the very same levels previously described and linear correlations were
363 calculated among the duration log ratios and the cost log ratios. A summary of this analysis is
364 presented in Table 5. Spearman's rho and Kendall's tau non-linear (rank) correlations were
365 also tested. However, they only very marginally improved the linear correlation results and
366 were considered not worth including as they did not seem to barely depart from the linear
367 case shown in Table 5.

368 **<Insert Table 5 here>**

369 Table 5 is divided in two major blocks. The upper block is devoted to activity-level
370 correlations. The lower block is reserved for project-level correlations. For each correlation it
371 has been specified how many datapoints were used (column labelled as n), Pearson's
372 correlation coefficient (R), the coefficient of determination (R^2), along with the gradient
373 (*slope* column) and *intercept* of the linear regression lines. Statistically highly significant
374 correlations have been marked with two asterisks (**) separately for R^2 tests (with the

375 Snedecor's F distribution) and *slope* tests (with the Student's T distribution). Significant
376 statistical correlations have been marked with a single asterisk (*).

377 In the case of activity-level correlations, almost all correlation values are significant.
378 This mean that values of R^2 are very unlikely to have happened by chance. This is not the
379 case at Project-level correlations where, apart from the level 0 of analysis (all 101 projects
380 grouped together), R^2 values have not been found to be statistically significant. This means
381 we cannot count on the reliability of project-level duration-cost correlations, hence they will
382 be ignored moving forward.

383 Correlations at activity-level do offer very interesting results. R and R^2 evidence weak
384 to moderate correlations (R^2 ranging between 0.10 and 0.62), but the slopes of such
385 correlations are rather close to 0.50 in some levels and almost all of them are significant
386 (marked with ** or *). More precisely, when there is no differentiation among activities
387 (level 0), the slope is as high as 0.704. This means that a 100% activity duration extension (in
388 log scale) would cause a 70.4% cost increment on that activity. This is quite a high gradient.

389 Differentiating by project type (level 1), the slopes become more informative.
390 *Building* and *civil engineering* projects boast a gradient close to 0.5, that is, every 100% of
391 duration increment is likely to cause a 50% of cost increment for that activity. For the other
392 two types of projects we have no statistically significant slopes, despite it seems clear that
393 *industrial* projects (probably due to the higher component of electromechanical equipment in
394 the project budget) have lower slopes. On the contrary, *Services* projects, being more labor
395 intensive, have higher slopes.

396 Results by project phase (level 2) seem more homogeneous. However, only the
397 *execution* activities' slope is statistically significant. This level of correlation seems to
398 replicate the results previously provided for level 0.

399 Results at level 3 are again not that heterogeneous and they all are statistically
400 significant. However, there is nothing remarkable that has not been highlighted before.

401 A last note concerns the regression line intercepts (last column in Table 5). As can be
402 seen, these values remain above 0.02 (in log scale) most of the time. That is approximately
403 equivalent to an intercept of 5% in natural scale, which means that, no matter whether
404 activity duration extensions are materialized or not, costs are likely to increase around 5% by
405 default. These values are in line with the log ratio cost *averages* found in Table 4.

406

407 **Discussion**

408 So far, almost all analyses have focused on individual activities. Yet, it is
409 acknowledged that the construction process is not an exact science and construction managers
410 are often ‘judged’ upon their capability to manage activity variability. Hence, the key concern
411 is the whole project suffering from delays and cost overruns, not just some of its activities. It
412 was proposed earlier that this is because activities suffer from variability (both positive and
413 negative), not because they are delayed on average. This section is devoted to analyze
414 whether this speculation seems acceptable.

415 Let us start by approaching the problem from a graphical perspective first. For that
416 purpose, a second dataset of 746 road construction projects from the Florida Department of
417 Transportation (USA) is used. Given the number of contracts, no descriptive table is included
418 in the paper, but the complete dataset can be found as *supplemental online material*. This
419 additional project dataset has been used here because they represent relatively similar
420 (homogeneous) contracts, from the same client, and during a short period of time. Arguably,
421 this is the closest to assuming that these projects are 746 different realizations (possible

422 outcomes) of the same generic type of project (in this case a road construction, that is, a *civil*
423 *engineering* project).

424 Figure 1 represents the distributions of the log deviation ratios for durations and costs
425 for the 746 contracts (using expressions (1) and (2) at project-level, not activity-level).

426 **<Insert Figure 1 here>**

427 Concerning project duration deviations (curve with black circles), it closely resembles
428 an extreme value distribution of maxima (both Fréchet and Gumbel fits have been provided
429 for comparison in black colors). This means that the merge event bias takes an important role
430 when determining the actual project duration. Results in natural scale are, in this occasion,
431 almost identical but they have not been provided to avoid curve cluttering.

432 Furthermore, it is worth noting that the average project duration extension is around
433 0.21 (in log scale). For *Civil Engineering* projects in Table 3, the average of the duration log
434 ratio was negative (-0.008). This means the activities from civil engineering projects ended
435 sooner than planned (on average). It is unlikely then, that the projects represented in Figure 1
436 could have ended later because a significant proportion of their activities ended late.

437 However, the activity duration variability (the standard deviation) was 0.20. In extreme value
438 theory, the mean of the highest order statistic distribution of a Normal distribution with three
439 or four draws is approximately one standard deviation. The Normal distribution represents
440 very well the distribution of the durations of each path (before they merge) (Ballesteros-
441 Pérez, 2017a). Therefore, the average of the duration distribution coincides very closely with
442 what is to be expected from the data from Table 3 for civil engineering projects ($0.21 \approx 0.20$).

443 Later it will be shown how more than three paths are quite common in civil engineering
444 construction schedules.

445 Concerning project costs (curve with grey crosses), the situation is very different. The
446 distribution of costs (log scale) resembles a Gamma distribution. It is worth noting that when
447 a random variable X follows a Pareto distribution with parameter λ , the logarithm of X
448 follows an Exponential distribution with the same parameter λ . This is relevant because the
449 costs of individual activities are well known to resemble a Pareto distribution in almost all
450 construction projects (Love et al. 2014; Love and Sing 2013). Hence, as cost ratios are being
451 processed here in log values, our distribution should also resemble an Exponential
452 distribution (continuous grey line in Figure 1). Additionally, as the sum of exponential
453 distributions is a Gamma distribution, that would offer some explanation, to why we are
454 observing a Gamma distribution (dashed grey line) fitting almost perfectly the log cost
455 deviations in Figure 1. In this case the exponential distribution also provides a good fit, but
456 that is not always the case in other construction project datasets.

457 Having approached the problem from a graphical and statistical perspective, it will be
458 addressed now from a topological perspective. Network topology describes the layout of
459 project schedules. The values of four representative topological indicators are displayed on
460 the last four columns of Table 1 for the 101 projects analyzed. Table 6 now shows the
461 average values of each topological indicator listed in Table 1, but categorized by Project type
462 (*building, civil engineering, industrial and services*), as well as for all projects together (last
463 row).

464 **<Insert Table 6 here>**

465 The Serial-Parallel (SP) indicator is probably the most relevant of the four indicators
466 for the purpose of this study. This indicator measures the closeness of a network to a serial or
467 parallel network. Namely, $SP = (m-1)/(n-1)$; where n is the total number of project activities
468 in a project schedule, and m is the number of activities in the path with a higher number of
469 activities (which may not necessarily be the longest in duration, as topological measures

470 ignore the activity durations). Hence, $SP=0$ means all activities are in parallel, whereas
471 $SP=100\%$ means all activities are in series. This indicator can also be considered as an
472 estimate of the amount of critical and non-critical activities in a network (Vanhoucke and
473 Vandevoorde 2009). Therefore, rounded up values of the inverse of the SP (that is $\lceil 1/SP \rceil$)
474 provide us with an estimate of the minimum number of paths of a project schedule. Values of
475 SP below 50% would mean that construction schedules have (approximately) at least three
476 paths. This agrees with what we appreciated in the black curve of Figure 1. *Industrial*
477 projects, despite having on average at least two paths, generally have a dominant one (which
478 condenses, on average, 55% of the activities). In *service* projects schedules there are at least
479 five paths (on average), as only 20% (a fifth) of the activities are critical.

480 Activity Distribution (AD) measures the distribution of project activities along the
481 levels of the project. In network topology, the number of project levels can be loosely defined
482 as the number of activities that are arrayed in parallel in a project schedule. Hence, AD
483 measures the width of the network. However, it is worth noting that activities arrayed in
484 parallel do not necessarily have to be executed simultaneously (because they may have different
485 time lags and/or activity durations). When $AD=0$ all levels contain a similar number of
486 activities and the number of activities is uniformly distributed over all levels. When
487 $AD=100\%$ there is one level with a maximal number of activities, and all other levels contain
488 a single activity. All four types of projects average AD values are close to 58% indicating
489 that the longest path has more activities than other paths, but still those other paths contain a
490 significant number of activities, that is, they can potentially cause project delays.

491 The Length of Arcs (LA) indicator measures the tightness of each precedence
492 relationship between two activities as the distance between two activities in the project
493 network. When $LA=0$ the network has many precedence relationships between two activities
494 on levels far from each other such that the activity can be shifted further in the network.

495 When LA=100%, many precedence relationships have a length of one, resulting in activities
496 with immediate successors on the next level of the network and with little freedom to shift.
497 Average LA values are much closer to 0 than to 100% (overall average of 14.1%). This
498 means that activities tend to have many predecessors (on average) from different levels
499 (paths), which would reinforce the merge event bias effect.

500 Finally, the Topological Float (TF) measures the degrees of freedom per activity as
501 the amount of slack or float an activity has. When TF=0 the network structure is 100% dense
502 and no activities can be shifted within its structure. When TF=100% the schedule consists of
503 a single chain of activities without topological float. The average TF indicator value of 40.5%
504 means that the average activity structure of construction projects is rather dense.

505 Therefore, the highlights of this brief topological analysis above for construction
506 projects are that: construction schedules are relatively dense (activity-wise), usually
507 composed of at least three major paths, and with activities whose predecessors usually come,
508 not just from activities located on the same path, but also from other paths. This means that
509 the merge event bias plays a very important role in construction schedules. And, precisely
510 thanks to the high level of duration variability existing at activity level, many delays are
511 expected to cumulate every time two or more paths merge into a single successor.

512 However, mergers are much more frequent towards the end of the project compared to
513 the earlier stages of execution. This is as, for any paths to close, they have to open first.
514 Therefore, it is not a surprise that many construction projects get off to a good start (on time
515 and on budget), but half way across their duration, (local) delays start being detected
516 (whenever two or more paths are merged into one). As delays emerge, the cost of activities
517 will also increase proportionally as the correlation between duration deviations and cost
518 deviations was quite substantial on average. As a result, it is not that surprising that projects
519 end later and cost more than initially anticipated.

520

521 **Conclusions**

522 The activity duration and cost variability of construction projects has been analyzed in
523 this research by different types of projects (*building, civil engineering, industrial and*
524 *services*), project phase (*planning, execution and closure*), and activity type (*auxiliary works,*
525 *substructure, superstructure, specialized works, and facilities*). Correlation factors between
526 activity duration deviations and activity cost deviations have also been studied under the
527 same activity categories. The research is novel because it describes the first four moments
528 (average, standard deviation, skewness and kurtosis) of how actual versus planned durations
529 and costs differ at activity level in construction projects. A set of 101 projects and 5289
530 activities, plus another set with 746 projects have been used.

531 The first contribution of this study is providing construction managers with a first, yet
532 rather complete, set of actual-vs-planned average activity durations and costs deviations with
533 application in multiple contexts (project types, execution phases and types of activity). From
534 now on, a construction manager will be able to more realistically (thus accurately) anticipate
535 how likely and how much the activities in the project schedule will vary, that is, last or cost
536 something different. This might potentially improve the quality and robustness of all
537 construction schedules, for example allowing them to feed more advanced (non-
538 deterministic) scheduling and simulation tools with more representative data. These
539 techniques generally need a substantial amount of information from previous similar projects
540 which is rarely available. With the set of moments provided here, these techniques will be
541 able to resort to average values for their activity durations and cost distribution parameters
542 depending on the type and/or execution phase of the project. These distributions will also be
543 able to assume non-independence between the stochastically-generated activity durations and
544 costs values (thanks to the set of duration-cost correlation values also published in this study).

545 This is expected to enhance future construction project monitoring and control, but also
546 actual project duration and cost forecasting accuracy.

547 However, the analysis developed has also provided some interesting insights from its
548 numerical perspective. One of the most relevant is that it has been shown that construction
549 activities do not end late on average. Instead, it is their high level of variability (around 60%
550 of its average duration) the key factor eventually causing project-level delays. Such high
551 levels of activity variability exacerbate the merge event bias, a phenomenon by which
552 whenever two or more schedule paths converge into a single one, the average completion
553 times exceed the maximum average path durations.

554 Actual activity costs, on the other hand, do tend to be higher than what was planned
555 (around 7%). This cannot be the result of price adjustments or inflation, as hardly any project
556 lasted longer than a year. Instead, the major project-level cost overruns are expected to occur
557 as a consequence of delayed start of activities located nearer the end of the project. This, as it
558 has been demonstrated how most duration-cost correlation factors range within 0.40 and 0.70.
559 The latter would cause that those activities that cannot start until their predecessors have
560 finished, start incurring in costs before their actual execution.

561 Many other interpretations can arise from the numerical results of the four moments
562 describing activity duration and cost variability that refer to specific types of projects, phases
563 of execution or activity types that have not been recounted here. The reader is invited to refer
564 to Tables 3 and 4 for such a purpose.

565 A limitation of this study is mostly connected to the composition and sample size of
566 the construction projects analyzed. 101 projects have been used here with a varied
567 composition. However, this sample size could have been bigger. It must be clarified, though,
568 that accessing actual duration and cost information is ontologically questionable and certainly

569 methodologically challenging. Companies are not open to share this information because it
570 would clearly indicate how competent and efficient their operations are. Under that
571 perspective, the current sample size probably seems satisfactory, at least for a first
572 representative analysis.

573 A second limitation arises from having removed at the outset the Unplanned but
574 Performed (UbP) and Planned but Unperformed (PbU) activities. This was necessary as the
575 ratios (either in natural or log scale) converged to infinity causing a distortion in the moments
576 calculation. However, we acknowledge that these activities can be found in almost all real
577 projects. Frequently, they are the consequence of scope changes, works reorganization or
578 changes in the available resources. Obviously, UbP and PbU activities add to the total project
579 variability (beyond the activity duration and cost variability analysed here). In our analysis,
580 though, there were only $279 \text{ UbP} + 129 \text{ PbU} = 408$ activities out of the initial 5,697 (7% in
581 total). Hence, while we believe the influence of UbP and PbU activities needs to be duly
582 investigated, our analysis (with 93% of the activities) can still be considered representative
583 enough to draw valid conclusions. Additionally, it is also expected that some degree of
584 cancellation will occur among those 7% of activities (as frequently new activities replace
585 others which are not eventually performed).

586 In the same vein, there are many potential future research continuations after this
587 piece of research. Again, this study might be extended to analyze other types of projects
588 and/or other more specific types of activities (maybe at trade-level: concrete, steel, asphalt,
589 earthworks, etc.). The network topologies for other types of projects may also be studied to
590 anticipate to what extent current levels of activity variability might impact their final
591 schedules. The statistical distribution of activity (duration and cost) variability may also be
592 analyzed. This was not possible at the general activity-level as discussed in this paper, but it
593 should be possible for activities at their trade level.

594 A last conclusion derived from this research is that activity duration variability is the
595 actual foe in project monitoring and control. This may not sound new to Lean Construction
596 researchers and practitioners. However, this research has provided compelling empirical
597 evidence suggesting that we do really need to start taking activity variability more seriously.
598 There is a need to develop more techniques that can effectively handle/restrain this
599 variability. Value stream mapping and Last planner have been some attempts to address this
600 problem, but more are needed. This will open the door to new and more effective approaches
601 for tackling the widespread phenomenon of construction projects ending late.

602

603 **Acknowledgements**

604 The first author acknowledges the Spanish Ministry of Science, Innovation and Universities
605 for his Ramon y Cajal contract (RYC-2017-22222) co-funded by the European Social Fund.
606 This work was also supported by the second author's "Estancias de movilidad en el
607 extranjero José Castillejo para jóvenes doctores, 2017 (grant ref. CAS17/00488)" and the
608 fourth author's "Estancias de profesores e investigadores senior en centros extranjeros
609 Salvador de Madariaga 2018 (grant ref. PRX18/00381)", both from the Spanish Ministry of
610 Science, Innovation and Universities. The first and last authors also acknowledge the help
611 received by the research group TEP-955 from the PAIDI (Junta de Andalucía, Spain).

612

613 **Data availability**

614 All data generated or analyzed during the study are included in the submitted article
615 or supplemental materials files.

616

617 **References**

- 618 AlSehaimi, A., and Koskela, L. (2008). "What Can be Learned from Studies on Delay in
619 Construction?" *Proceedings for the 16th Annual Conference of the International Group*
620 *for Lean Construction*, 95–106.
- 621 Ballesteros-Pérez, P. (2017a). "M-PERT: Manual Project-Duration Estimation Technique for
622 Teaching Scheduling Basics." *Journal of Construction Engineering and Management*,
623 143(9).
- 624 Ballesteros-Pérez, P. (2017b). "Modelling the boundaries of project fast-tracking."
625 *Automation in Construction*, 84(December), 231–241.
- 626 Ballesteros-Pérez, P., Campo-Hitschfeld, M. L., González-Naranjo, M. A., and González-
627 Cruz, M. C. (2015). "Climate and construction delays: case study in Chile."
628 *Engineering, Construction and Architectural Management*, 22(6), 596–621.
- 629 Ballesteros-Pérez, P., Rojas-Céspedes, Y. A., Hughes, W., Kabiri, S., Pellicer, E., Mora-
630 Melià, D., and del Campo-Hitschfeld, M. L. (2017). "Weather-wise. A weather-aware
631 planning tool for improving construction productivity and dealing with claims."
632 *Automation in Construction*, 84(December), 81–95.
- 633 Ballesteros-Pérez, P., Larsen, G. D., and González-Cruz, M. C. (2018a). "Do projects really
634 end late? On the shortcomings of the classical scheduling techniques." *Journal of*
635 *Technology and Science Education*, 8(1), 86–102.
- 636 Ballesteros-pérez, P., Smith, S. T., Lloyd-papworth, J. G., and Cooke, P. (2018b).
637 "Incorporating the effect of weather in construction scheduling and management with
638 sine wave curves : application in the United Kingdom." *Construction Management and*
639 *Economics*, 36(12), 666–682.
- 640 Banerjee, A., and Paul, A. (2008). "On path correlation and PERT bias." *European Journal*
641 *of Operational Research*, 189(3), 1208–1216.
- 642 Batselier, J., and Vanhoucke, M. (2015). "Construction and evaluation framework for a real-
643 life project database." *International Journal of Project Management*, 33(3), 697–710.
- 644 Bein, W. W., Kamburowski, J., and Stallmann, M. F. M. (1992). "Optimal Reduction of
645 Two-Terminal Directed Acyclic Graphs." *SIAM Journal on Computing*, 21(6), 1112–
646 1129.
- 647 Cho, S. (2009). "A linear Bayesian stochastic approximation to update project duration
648 estimates." *European Journal of Operational Research*, 196(2), 585–593.
- 649 Chudley, R., and Greeno, R. (2016). *Building Construction Handbook*. Routledge; 11th
650 edition. ISBN-10: 113890709X
- 651 Davies, E. M. (1973). "An Experimental Investigation of Resource Allocation in
652 Multiactivity Projects." *Operational Research Quarterly (1970-1977)*, *Journal of the*
653 *Operational Research Society*, 24(4), 587.
- 654 Dodin, B., and Sirvanci, M. (1990). "Stochastic networks and the extreme value distribution."
655 *Computers & Operations Research*, 17(4), 397–409.
- 656 Elmaghraby, S. E. (2000). "On criticality and sensitivity in activity networks." *European*
657 *Journal of Operational Research*, 127(2), 220–238.
- 658 Elmaghraby, S. E., Fathi, Y., and Taner, M. R. (1999). "On the sensitivity of project
659 variability to activity mean duration." *International Journal of Production Economics*,
660 62(3), 219–232.

- 661 George, D., and Mallery, P. (2010). *SPSS for Windows Step-by-Step: A Simple Guide and*
662 *Reference, 17.0 Update.* (Pearson, ed.), *Pearson*, Boston.
- 663 Hajdu, M., and Bokor, O. (2014). “The Effects of Different Activity Distributions on Project
664 Duration in PERT Networks.” *Procedia - Social and Behavioral Sciences*, 119, 766–
665 775.
- 666 Hajdu, M., and Bokor, O. (2016). “Sensitivity analysis in PERT networks : Does activity
667 duration distribution matter ?” *Automation in Construction*, Elsevier B.V., 65, 1–8.
- 668 Hamzah, N., Khoiry, M. A., Arshad, I., Tawil, N. M., and Che Ani, A. I. (2011). “Cause of
669 Construction Delay - Theoretical Framework.” *Procedia Engineering*, 20, 490–495.
- 670 Keane, P. J., and Caletka, A. F. (2008). *Delay analysis in construction contracts.* *Wiley-*
671 *Blackwell*, Oxford, U.K.
- 672 Khamooshi, H., and Cioffi, D. F. (2013). “Uncertainty in Task Duration and Cost Estimates:
673 Fusion of Probabilistic Forecasts and Deterministic Scheduling.” *Journal of*
674 *Construction Engineering and Management*, 139(5), 488–497.
- 675 Love, P. E. D., Lopez, R., Kim, J. T., and Kim, M. J. (2014). “Probabilistic Assessment of
676 Design Error Costs.” *Journal of Performance of Constructed Facilities*, 28(3), 518–527.
- 677 Love, P. E. D., and Sing, C.-P. (2013). “Determining the probability distribution of rework
678 costs in construction and engineering projects.” *Structure and Infrastructure*
679 *Engineering*, 9(11), 1136–1148.
- 680 MacCrimmon, K. R., and Ryavec, C. A. (1964). “An Analytical Study of the PERT
681 Assumptions.” *Operations Research*, 12(1), 16–37.
- 682 Mahamid, I., Bruland, A., and Dmaid, N. (2012). “Causes of Delay in Road Construction
683 Projects.” *Journal of Management in Engineering*, 28(3), 300–310.
- 684 Mastor, A. A. (1970). “An Experimental Investigation and Comparative Evaluation of
685 Production Line Balancing Techniques.” *Management Science*, 16(11), 728–746.
- 686 Ogunlana, S. O., Promkuntong, K., and Jearkijrm, V. (1996). “Construction delays in a fast-
687 growing economy: Comparing Thailand with other economies.” *International Journal of*
688 *Project Management*, 14(1), 37–45.
- 689 OR-AS.be. (2018). “Real Project Data.” Last accessed on 23th April 2019:
690 <http://www.projectmanagement.ugent.be/research/data/realdata>
- 691 Orangi, A., Palaneeswaran, E., and Wilson, J. (2011). “Exploring Delays in Victoria-Based
692 Australian Pipeline Projects.” *Procedia Engineering*, 14, 874–881.
- 693 Pascoe, T. (1966). “Allocation of resources: CPM.” *Revue Française de Recherche*
694 *Opérationnelle*, 38, 31–38.
- 695 Project Management Institute. (2017). *PMBOK® Guide – Sixth Edition.* Newtown Square,
696 Pa. ISBN-10: 1628251840
- 697 Senouci, A., Ismail, A., and Eldin, N. (2016). “Time Delay and Cost Overrun in Qatari Public
698 Construction Projects.” *Procedia Engineering*, 164, 368–375.
- 699 Trietsch, D., Mazmany, L., Gevorgyan, L., and Baker, K. R. (2012). “Modeling activity
700 times by the Parkinson distribution with a lognormal core: Theory and validation.”
701 *European Journal of Operational Research*, 216(2), 386–396.
- 702 Valadares Tavares, L., Antunes Ferreira, J., and Silva Coelho, J. (1999). “The risk of delay of
703 a project in terms of the morphology of its network.” *European Journal of Operational*
704 *Research*, 119(2), 510–537.

- 705 Vanhoucke, M. (2008). “Measuring time using novel earned value management metrics.”
706 *Proceedings of the 22nd IPMA World Congress (Rome)*, vol 1, 99–103.
- 707 Vanhoucke, M. (2012). *Project Management with Dynamic Scheduling*. Springer Berlin
708 Heidelberg, Berlin.
- 709 Vanhoucke, M., Coelho, J., and Batselier, J. (2016). “An Overview of Project Data for
710 Integrated Project Management and Control.” *The Journal of Modern Project*
711 *Management*, 3(3), 6–21.
- 712 Vanhoucke, M., and Vandevorde, S. (2009). “Forecasting a project’s duration under various
713 topological structures.” *The Measurable News*, (Spring), 26–30.

Project ID	Project Name	Project Type	Planned Cost (€)	Actual Cost (€)	Planned Dur. (d)	Actual Dur. (d)	N° activ.	SP (%)	AD (%)	LA (%)	TF (%)
C2011-05	Telecom System Agnes	Service	180,485.27	180,485.27	43	53	20	60	58	38	9
C2011-07	Patient Transport System	Service	180,759.44	191,065.06	389	444	49	70	70	7	8
C2011-10	Building a House	Building	484,398.41	494,947.71	195	203	32	51	47	27	10
C2011-12	Claeys-Verhelst Premises	Building	3,027,133.19	3,102,395.91	443	453	49	41	50	5	43
C2011-13	Wind Farm	Civil Eng.	21,369,835.51	26,077,764.74	525	600	107	27	36	0	48
C2012-13	Pumping Station Jabbeke	Industrial	336,410.15	350,511.31	125	140	74	64	59	3	27
C2012-15	The Master Project	Service	185,472.45	185,113.10	32	32	121	17	66	0	84
C2012-17	Building a Dream	Building	241,015.00	314,856.14	145	204	33	65	61	35	19
C2013-01	Wiedauwkaai Fenders	Civil Eng.	1,069,532.42	1,314,584.58	152	152	39	48	45	0	68
C2013-02	Sewage Plant Hove	Civil Eng.	1,236,603.66	1,146,444.38	403	408	175	12	38	0	62
C2013-03	Brussels Finance Tower	Building	15,440,865.89	16,338,027.20	425	426	55	3	82	0	87
C2013-04	Kitchen Tower Anderlecht	Building	2,113,684.00	2,512,524.00	333	453	244	47	59	0	63
C2013-05	PET Packaging	Service	874,554.28	874,554.28	521	632	28	14	69	0	80
C2013-06	Govmnt. Office Building	Building	19,429,810.51	21,546,846.18	352	344	275	10	36	0	34
C2013-07	Family Residence	Building	180,476.47	175,030.65	170	174	46	40	44	3	25
C2013-08	Timber House	Building	501,029.51	576,624.05	216	235	41	29	42	0	47
C2013-09	Urban Develop.Project	Civil Eng.	1,537,398.51	1,696,971.79	291	360	71	34	51	6	16
C2013-10	Town Square	Civil Eng.	11,421,890.36	15,218,926.38	786	785	186	18	36	0	62
C2013-11	Recreation Complex	Building	5,480,518.91	5,451,028.00	359	277	159	27	44	0	32
C2013-12	Young Cattle Barn	Building	818,439.99	879,853.17	115	188	27	64	77	6	54
C2013-13	Office Finish. Works (1)	Building	1,118,496.59	955,929.22	236	217	11	20	49	33	6
C2013-14	Office Finish. Works (2)	Building	85,847.89	75,468.30	80	88	9	62	80	66	47
C2013-15	Office Finish. Works (3)	Building	341,468.11	308,343.78	171	115	17	25	43	21	35
C2013-16	Office Finish. Works (4)	Building	248,203.92	198,567.00	196	108	7	33	62	0	75
C2013-17	Office Finish. Works (5)	Building	244,205.40	203,605.97	161	107	23	36	38	20	32
C2014-01	Mixed-use Building	Building	38,697,822.73	39,777,643.30	474	448	41	50	38	3	49
C2014-02	Playing Cards	Industrial	191,492.70	190,266.50	124	146	21	81	94	0	14
C2014-03	Organizational Develop.	Service	43,170.15	83,712.15	229	260	112	9	31	0	36
C2014-04	Compres. Station Zelzate	Industrial	62,385,597.58	65,526,930.04	522	844	24	95	100	0	100
C2014-05	Apartment Building (1)	Building	532,410.29	591,410.53	228	274	25	58	71	35	18
C2014-06	Apartment Building (2)	Building	3,486,375.47	3,599,114.11	547	611	29	57	75	46	15
C2014-07	Apartment Building (3)	Building	1,102,536.78	1,289,696.78	353	404	25	58	71	35	18
C2014-08	Apartment Building (4)	Building	1,992,222.09	2,380,299.86	233	275	39	44	29	11	14
C2015-01	Young Cattle Barn (2)	Building	612,769.44	646,473.65	131	210	27	57	73	0	46
C2015-02	Railway Station (1)	Civil Eng.	1,121,316.94	967,988.79	417	501	216	8	66	1	80
C2015-03	Industrial Complex (1)	Building	2,244,090.74	1,868,796.28	257	278	135	16	43	0	58
C2015-04	Apartment Building (5)	Building	2,750,938.00	2,590,796.73	160	205	56	27	37	0	57
C2015-06	Family Residence (2)	Building	143,673.20	186,107.00	260	290	184	18	0	30	38
C2015-07	Industrial Complex (2)	Building	5,999,600.00	5,414,544.00	297	313	138	27	38	0	49
C2015-08	Garden Center	Building	467,297.21	461,900.17	191	186	186	14	52	0	79
C2015-09	Railway Station (2)	Civil Eng.	1,457,424.00	2,145,682.26	354	569	340	4	48	0	75
C2015-10	Tax Return System (1)	Service	18,990.00	8,010.00	85	85	15	10	82	23	21
C2015-11	Staff Authoriz. System	Service	14,400.00	9,105.00	55	55	7	25	66	0	52
C2015-12	Premium Payment System	Service	132,570.00	58,410.00	184	184	35	19	63	9	61
C2015-13	Broker Acc.Conv. System	Service	12,735.00	9,990.00	117	117	16	19	60	7	51
C2015-14	Sup. Pensions Database	Service	34,260.00	18,285.00	124	124	17	17	55	3	50
C2015-15	FACTA System	Service	11,700.00	7,035.00	57	57	13	22	57	8	18
C2015-16	Generic Doc. Output Syst.	Service	64,620.00	64,125.00	270	270	22	10	61	12	26
C2015-17	Insurance Bundling Syst.	Service	281,430.00	281,070.00	208	236	86	6	77	8	41
C2015-18	Tax Return System (2)	Service	39,450.00	25,380.00	128	128	15	10	66	16	11
C2015-19	Receipt Numb. System	Service	43,800.00	37,530.00	182	182	20	21	46	8	31
C2015-20	Policy Numbering System	Service	12,645.00	11,100.00	171	161	6	20	62	20	13

C2015-21	Investment Product (1)	Service	4,020.00	3,240.00	37	37	12	18	35	2	36
C2015-22	Risk Profile Questionnaire	Service	29,880.00	17,400.00	151	151	22	16	70	9	40
C2015-23	Investment Product (2)	Industrial	46,920.00	32,805.00	122	120	33	17	53	5	39
C2015-24	CRM System	Service	44,130.00	36,870.00	233	233	21	7	59	7	29
C2015-25	Beer Tasting	Service	1,210.00	1,780.00	14	14	18	16	40	21	19
C2015-26	Debt Collection System	Service	458,112.37	512,546.15	148	154	214	9	43	0	61
C2015-27	Railway Station Antwerp	Building	22,703.52	25,313.12	68	81	18	23	40	-2	54
C2015-28	Web. Tennis Vlaanderen	Service	219,275.00	382,475.00	201	212	20	15	54	0	67
C2015-29	Fire Station	Building	1,874,496.82	1,887,087.25	284	298	204	48	34	0	41
C2015-30	Social Apts. Ypres (1)	Building	440,940.89	440,940.89	244	254	40	25	51	-1	76
C2015-31	Social Apts Ypres (2)	Building	1,310,723.46	1,282,185.98	271	364	29	32	49	23	43
C2015-32	Social Apts Ypres (3)	Building	2,509,031.42	2,509,031.42	358	265	48	38	63	3	59
C2015-33	IJzertoren Memor. Square	Civil Eng.	214,417.71	224,789.67	50	94	12	63	57	0	14
C2015-34	Roadworks Poperinge	Civil Eng.	511,325.86	440,394.16	120	193	13	91	99	0	18
C2015-35	Retirement Apartments	Building	14,956,314.25	16,068,878.30	850	951	11	48	57	21	35
C2016-01	Railway Bridge (1)	Civil Eng.	671,383.50	703,703.50	225	274	26	51	71	0	86
C2016-02	Railway Bridge (2)	Civil Eng.	962,181.56	972,341.56	229	239	23	63	71	0	82
C2016-03	Railway Bridge (3)	Civil Eng.	926,888.01	910,728.01	203	220	25	16	37	0	56
C2016-04	Railway Bridge (4)	Civil Eng.	906,253.87	906,253.87	248	242	26	64	62	0	71
C2016-05	Railway Bridge (5)	Civil Eng.	832,497.46	832,497.46	195	197	32	77	74	0	51
C2016-06	Defense Building	Service	4,331,260.49	4,331,260.49	252	232	96	14	55	0	76
C2016-07	Shop. Village Walkways	Civil Eng.	930,179.09	932,757.25	224	316	110	95	98	0	99
C2016-08	SCM System	Service	375,253.34	438,741.66	725	725	99	49	59	8	52
C2016-09	Data Loss Prevent. System	Service	584,951.77	1,425,155.96	195	189	113	10	36	1	51
C2016-10	Biofuel Refinery	Industrial	14,362,625.00	14,466,100.00	360	375	23	18	22	6	21
C2016-11	Residential House (1)	Building	162,472.00	163,189.00	241	254	55	57	77	52	16
C2016-12	Residential House (2)	Building	222,858.00	226,285.00	291	291	59	56	72	50	19
C2016-13	Residential House (3)	Building	367,952.00	379,300.00	306	330	51	64	81	54	14
C2016-14	Residential House (4)	Building	218,366.00	222,021.78	321	320	48	68	78	42	10
C2016-15	Resid. House Struct. Work	Building	95,694.00	100,763.00	126	130	13	66	75	100	0
C2016-16	Resid. Finish. Works (1)	Building	54,577.76	64,526.76	90	90	24	69	68	50	28
C2016-17	Resid. Finish. Works (2)	Building	54,703.17	64,580.17	86	86	24	69	68	50	28
C2016-18	Resid. Finish. Works (3)	Building	51,115.52	60,829.52	91	91	25	66	62	27	31
C2016-19	Resid. Finish. Works (4)	Building	51,303.38	53,351.38	91	91	25	66	62	27	31
C2016-20	Resid. Finish. Works (5)	Building	52,021.28	53,783.28	91	91	25	66	62	27	31
C2016-21	Resid. Finish. Works (6)	Building	54,324.22	54,996.22	101	101	24	69	68	50	28
C2016-22	Resid. Finish. Works (7)	Building	56,969.40	57,822.40	101	101	24	69	68	50	28
C2016-23	Resid. Finish. Works (8)	Building	56,182.71	56,645.71	101	101	24	69	68	50	28
C2016-24	Resid. Finish. Works (9)	Building	52,262.83	53,176.83	101	101	24	69	68	50	28
C2016-25	Resid. Finish. Works (10)	Building	54,580.33	56,748.33	91	91	24	69	68	50	28
C2016-26	Resid. Finish. Works (11)	Building	51,286.24	53,319.24	91	91	24	69	68	50	28
C2016-27	Apt. Build. Foundat. (1)	Building	813,663.06	879,701.06	78	88	16	66	59	0	48
C2016-28	Apt. Struct. Work (1)	Building	569,177.85	586,086.85	71	79	19	55	29	0	30
C2016-29	Apt. Struct. Work (2)	Building	1,797,873.62	1,860,330.62	129	148	19	72	69	0	35
C2016-30	Apt. Struct. Work (3)	Building	1,319,736.29	1,353,361.29	85	96	23	81	83	0	31
C2016-31	Apt. Struct. Work (1)	Building	488,936.00	498,473.00	105	117	23	31	40	0	11
C2016-32	Apt. Struct. Work (2)	Building	477,381.00	496,991.00	89	97	22	52	72	0	27
C2016-33	Apt. Struct. Work (3)	Building	377,282.00	394,829.00	116	129	23	50	72	0	30
C2016-34	Apt. Struct. Work (4)	Building	362,476.00	383,871.00	83	92	23	40	43	0	26
Avg.			2,647,861.81	2,837,446.83	221	240	$\Sigma=5,697$	41.0	58.2	14.1	40.5

Table 1. First projects dataset summary

Level 0 (All activities*)

N° activities		
Planned & Performed	Unplanned but Performed	Planned but not performed
5289	279	129

Level 1 (by Project type*)

Project Type	n	N° activities			Actual Cost (10 ³ €)		Actual Dur. (days)	
		Planned & Performed	Unplanned but Performed	Planned but not performed	Avg.	SD	Avg.	SD
Building	56	2894	18	12	48.88	267.20	11.35	29.78
Civil Eng.	15	1092	250	59	40.43	161.26	12.92	15.92
Industrial	5	170	0	5	473.92	1225.85	21.60	48.60
Services	25	1133	11	53	8.03	31.15	11.13	31.48
Sum	101	5289	279	129				

Level 2 (by Project phase*)

Proj. type > Project phase V	N° activities							
	Building		Civil Engineering		Industrial		Service	
	Plan. & Perform.	Unplan. but Perform.	Plan. & Perform.	Unplan. but Perform.	Plan. & Perform.	Unplan. but Perform.	Plan. & Perform.	Unplan. but Perform.
Planning	49	0	38	0	10	0	81	0
Execution	2810	18	1034	250	154	0	990	11
Closure	35	0	20	0	6	0	62	0

Level 3 (by Activity type **&***)

Project type > Activity type V	N° activities					
	Planned & Performed	Building		Planned but not perform.	Civil Engineering	
Unplan. but Performed		Unplan. but Performed	Planned & Performed		Unplan. but Performed	Planned but not perform.
Auxiliary works	139	1	0	207	27	9
Substructure	171	2	0	229	11	4
Superstructure	654	1	0	257	104	20
Specialized works	1272	11	10	264	88	25
Facilities	574	3	2	77	20	1

*Only Planned & Performed activities are used for later analyses

** Only for 'Execution' activities from Building and Civil Engineering projects

Table 2. Summary of activities analyzed

Level 0 (All Activities)					Level 1 (by Project Type)					Level 2 (by Project phase)					Level 3 (by Activity type)							
n	Avg	SD	Skew	Kurt	Type	n	Avg	SD	Skew	Kurt	Phase	n	Avg	SD	Skew	Kurt	Type	n	Avg	SD	Skew	Kurt
5289	0.010 (1.023)	0.19 (1.56)	0.91	9.90	Building	2894	0.004 (1.009)	0.15 (1.43)	-0.36	9.88	Planning	49	0.035 (1.083)	0.21 (1.62)	1.51	8.13	<i>(insufficient data sample)</i>					
											Execution	2810	0.003 (1.007)	0.15 (1.42)	-0.46	9.92	Auxiliary Works	139	0.017 (1.040)	0.16 (1.46)	0.36	5.54
																	Substructure	171	0.035 (1.083)	0.14 (1.39)	1.89	8.38
																	Superstructure	654	-0.018 (0.960)	0.16 (1.45)	-0.98	9.60
																	Specialized Works	1272	0.004 (1.010)	0.15 (1.42)	-0.87	10.02
											Closure	35	0.022 (1.052)	0.16 (1.45)	1.33	3.91	<i>(insufficient data sample)</i>					
					Civil Eng.	1092	-0.008 (0.982)	0.20 (1.58)	0.53	9.61	Planning	38	0.052 (1.126)	0.18 (1.53)	2.38	12.88	<i>(insufficient data sample)</i>					
											Execution	1034	-0.010 (0.977)	0.20 (1.58)	0.49	9.36	Auxiliary Works	207	0.013 (1.030)	0.13 (1.36)	1.75	9.43
																	Substructure	229	-0.005 (0.990)	0.18 (1.53)	-0.39	8.57
																	Superstructure	257	-0.030 (0.934)	0.20 (1.57)	0.90	10.73
																	Specialized Works	264	-0.012 (0.972)	0.24 (1.72)	0.28	5.81
											Closure	20	-0.011 (0.975)	0.05 (1.12)	-4.47	20.00	<i>(insufficient data sample)</i>					
					Industrial	170	-0.010 (0.977)	0.22 (1.65)	-0.76	3.37	Planning	10	0.001 (1.003)	0.05 (1.12)	0.43	4.59	<i>(insufficient data sample)</i>					
											Execution	154	-0.009 (0.981)	0.23 (1.68)	-0.76	3.17	<i>(insufficient data sample)</i>					
											Closure	6	-0.090 (0.813)	0.25 (1.77)	0.04	0.81	<i>(insufficient data sample)</i>					
					Services	1133	0.045 (1.110)	0.26 (1.83)	1.53	5.87	Planning	81	0.055 (1.134)	0.22 (1.67)	1.05	3.55	<i>(insufficient data sample)</i>					
											Execution	990	0.048 (1.118)	0.27 (1.86)	1.58	5.70	<i>(insufficient data sample)</i>					
											Closure	62	-0.014 (0.969)	0.19 (1.54)	-1.36	7.74	<i>(insufficient data sample)</i>					

Table 3. Activity actual/planned duration log ratios (natural values stated between parentheses)

Level 0 (All Activities)					Level 1 (by Project Type)					Level 2 (by Project phase)					Level 3 (by Activity type)															
n	Avg	SD	Skew	Kurt	Type	n	Avg	SD	Skew	Kurt	Phase	n	Avg	SD	Skew	Kurt	Type	n	Avg	SD	Skew	Kurt								
5289	0.031 (1.074)	0.25 (1.78)	2.49	15.56	Building	2894	0.015 (1.035)	0.16 (1.46)	2.02	25.27	Planning	49	-0.002 (0.996)	0.19 (1.56)	-1.66	10.45	<i>(insufficient data sample)</i>													
											Execution	2810	0.015 (1.035)	0.16 (1.46)	2.12	25.91	Aux. Works	139	0.027 (1.065)	0.11 (1.29)	2.73	14.82								
																	Substruct.	171	0.014 (1.034)	0.10 (1.26)	-0.21	8.37								
																	Superstruct.	654	0.010 (1.023)	0.10 (1.26)	0.01	10.14								
																	Spec. Works	1272	0.014 (1.034)	0.21 (1.62)	2.12	18.98								
																	Facilities	574	0.020 (1.046)	0.12 (1.33)	0.53	13.85								
																	Closure	35	0.041 (1.098)	0.16 (1.43)	2.01	6.60	<i>(insufficient data sample)</i>							
																												<i>(insufficient data sample)</i>		
										Civil Eng.	1092	0.057 (1.139)	0.30 (2.01)	1.78	6.28	Planning	38	0.322 (2.099)	0.43 (2.71)	0.99	-0.75	<i>(insufficient data sample)</i>								
					Execution	1034	0.048 (1.116)	0.30 (1.98)	1.77							6.87	Aux. Works	207	0.059 (1.147)	0.32 (2.08)	2.89	11.31								
																	Substruct.	229	0.057 (1.140)	0.30 (1.98)	0.63	2.26								
																	Superstruct.	257	0.057 (1.141)	0.32 (2.11)	1.40	3.11								
										Industrial	170	-0.011 (0.975)	0.20 (1.59)	-2.05	11.12	Planning	10	0.02 (1.046)	0.05 (1.03)	0.74	0.71	<i>(insufficient data sample)</i>								
					Execution	154	-0.004 (0.99)	0.18 (1.50)	-1.65							12.65	<i>(insufficient data sample)</i>													
					Closure	6	-0.27 (0.536)	0.67 (4.63)	-0.12							-2.71	<i>(insufficient data sample)</i>													
					Services	1133	0.052 (1.128)	0.36 (2.29)	2.25	9.01	Planning	81	0.021 (1.05)	0.18 (1.53)	0.40	3.89	<i>(insufficient data sample)</i>													
Execution	990	0.059 (1.145)	0.37 (2.37)	2.23							8.27	<i>(insufficient data sample)</i>																		
Closure	62	-0.007 (0.985)	0.29 (1.93)	1.28							13.16	<i>(insufficient data sample)</i>																		

Table 4. Activity actual/planned costs log ratios (natural values stated between parentheses)

Group of analysis		n	R	R ²	Slope	Intercept
<i>Activity-level (duration-cost correlations)</i>						
Level 0	All activities	5289	0.55	0.30**	0.704**	0.024
Level 1	Building	2894	0.46	0.21**	0.488**	0.013
	Civil Engineering	1092	0.33	0.11**	0.502**	0.061
	Industrial	170	0.11	0.01	0.106	-0.010
	Services	1133	0.79	0.62**	1.074	0.004
Level 2	Planning	178	0.38	0.15**	0.534	0.055
	Execution	4988	0.55	0.30**	0.706**	0.024
	Closure	123	0.60	0.36**	0.534	0.055
Level3	Auxiliary Works	349	0.34	0.12**	0.601*	0.037
	Substructure	400	0.25	0.06**	0.343*	0.035
	Superstructure	912	0.32	0.10**	0.289**	-0.028
	Specialized Works	1609	0.44	0.20**	0.566*	0.013
	Facilities	654	0.53	0.28**	0.522**	0.021
<i>Project-level (duration-cost correlations)</i>						
	All Projects	101	0.22	0.05*	0.156*	0.029
	Building	56	0.52	0.27	0.957	0.006
	Civil Engineering	15	0.01	0.00	0.017	0.080
	Industrial	5	0.56	0.31	0.629	0.083
	Services	25	0.23	0.05	0.039	0.014
	Road projects (Figure 1)	746	0.34	0.11	0.108	0.016

**Snedecor's F test (for R²) or student's T test (for slopes) significant at $\alpha < 0.001$

* Snedecor's F test (for R²) or student's T test (for slopes) significant at $\alpha < 0.05$

Table 5. Duration vs Cost (log ratios) linear correlations

Project type	n	SP (%)	AD (%)	LA (%)	TF (%)
Building	56	48.2	57.4	21.4	35.2
Civil Eng.	15	44.7	59.3	0.5	44.7
Industrial	5	55.0	65.6	2.8	55.0
Service	25	20.1	57.6	8.3	20.1
All	101	41.0	58.2	14.1	40.5

Table 6. Average network topological values by project type

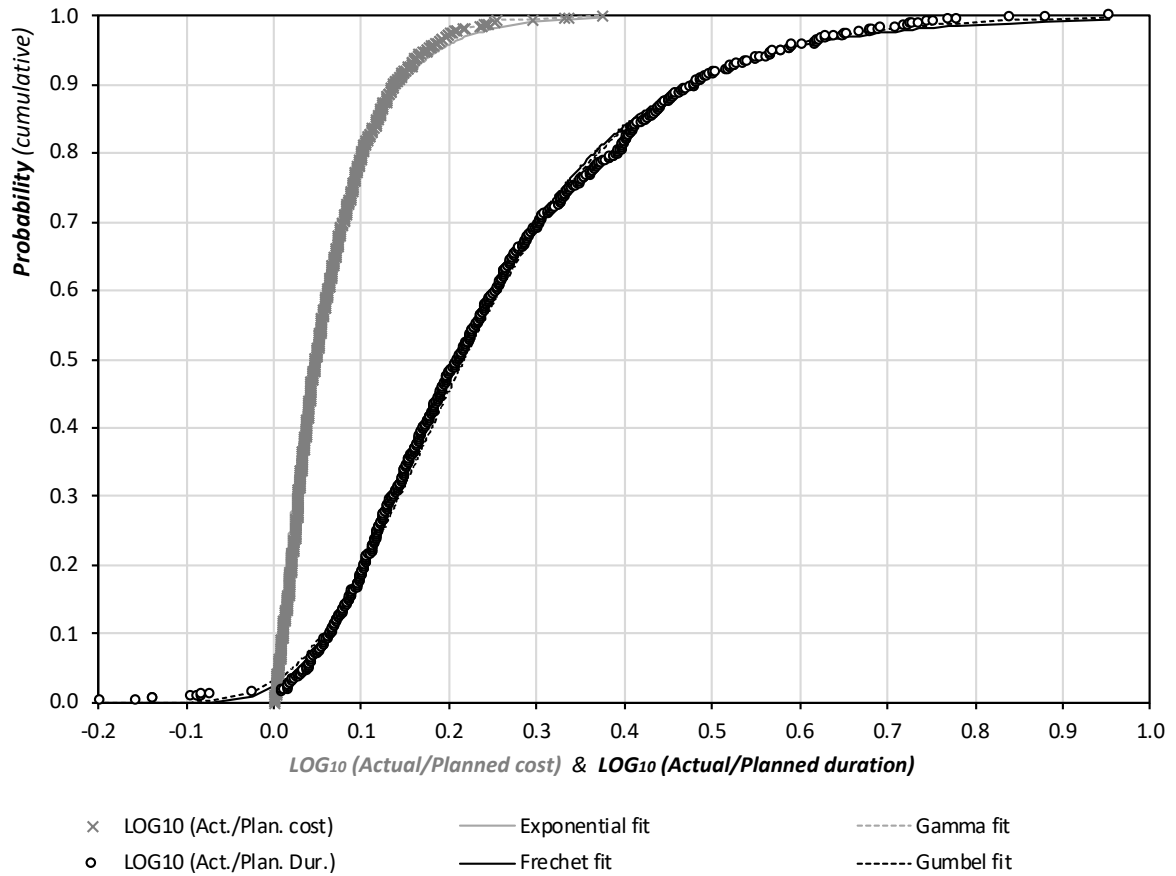


Fig 1. Duration and Cost overrun probability distribution of 746 road construction projects from the Florida Department of Transportation