MIT Open Access Articles

Inter-phase feedbacks in construction projects

Inter-phase Feedbacks in Construction Projects

Kiavash Parvan[a], Hazhir Rahmandad[b], Ali Haghani[c]

(a) PhD in project management program, Civil and Environmental Engineering Department, University of Maryland, kparvan@umd.edu

(b) Associate professor, Industrial and Systems Engineering Department, Virginia Polytechnic Institute and State University, hazhir@vt.edu

(c) Professor, Civil and Environmental Engineering Department, University of Maryland, haghani@umd.edu

Abstract

Understanding diverse performance trajectories of projects is of interest to operations researchers and practitioners. Interactions between multiple phases of a project are commonly assumed to be important in project dynamics, yet the strength of these feedback mechanisms has not been rigorously evaluated. In this study we use data from 15 construction projects to estimate the feedbacks between design and construction phases. The estimated factors reveal that undiscovered design rework diminishes construction quality and production rate significantly and construction completion speeds up the detection of undiscovered design rework. Together these feedbacks can explain as much as 20% of variability in overall project costs. Comparison of model predictions with a separate set of 15 projects shows good predictive power for cost and schedule outcomes and their uncertainty. The estimation and prediction framework offers a template for using data from multiple cases to estimate both case-specific and industry-wide parameters of project models, and for leveraging those estimates for project planning.

1- Introduction

Projects are critical to how modern organizations structure work. Moreover, faster product life cycles and increasingly global supply chains require firms to organize many steps of once routine operations, such as manufacturing and production processes, as projects (Gunasekaran and Ngai 2012). Therefore project management is increasingly an important part of operations management research and practice. Related operations literature has largely focused on designing algorithms for efficiently planning the execution of a project and offering decision support based on this algorithmic framework (Tavares 2002). Besides this academic literature, practitioner knowledge communities, such as the Project Management Institute, have elicited and codified best practices and offer various training and certification options to individuals and organizations (Williams 2005).

Despite the importance of projects and the management tools designed and applied by the research and practitioner communities, many projects fall short of their targets. From software (Moløkken and Jørgensen 2003) to construction (Mansfield, Ugwu and Doran 1994), infrastructure (Flyvbjerg, Holm and Buhl 2003), and military applications (Drezner, Jarvaise, Hess *et al.* 1993), delays, cost overruns, and quality problems have plagued many projects. For example, Standish Group's biennial surveys of the IT industry have found significant cost and budget overruns and cancellations in the majority of surveyed projects and a Procter & Gamble survey found that 15% of

authorized projects cost over 50% more than the original budget (Scott-Young and Samson 2008). Similarly, the U.S. General Accounting Office found cost overruns between 40% and 400% in a sample of 20 large infrastructure projects across 17 states (General Accounting Office 2002). In fact, disagreements over projects' fates instigate many legal disputes (Callahan, Bramble and Lurie 1990), some of them in the billions of dollars, and may even play a role in national politics (Pear, LaFraniere and Austen 2013).

Significant delays and cost overruns have motivated a more critical look at the assumptions that underlie operations research on projects. Specifically, the conventional project management paradigm is based on the decomposition of total project work into smaller, sequentially dependent tasks and algorithmic planning of the optimal sequencing and resource allocation within a project (Pollack 2007). However, there is an increasing realization that projects include many uncertainties and structural, dynamic, and socio-political complexities (Checkland and Winter 2006; Winter 2006; Geraldi, Maylor and Williams 2011). These complexities require a more systemic approach that is able to incorporate organizational and psychological antecedents to project performance (Hong, Nahm and Doll 2004; Bendoly and Swink 2007; Hagen and Park 2013) as well as feedback loops and nonlinearities that may trigger unintended consequences of the conventional approach to project planning (Ackermann, Eden and Williams 1997; Williams 2005; Lyneis and Ford 2007; Mingers and White 2010).

In response, research into understanding the root causes of project challenges has followed two complementary directions. Several empirical studies in this domain have used surveys of clients, contractors, and design personnel to assess the magnitude of cost

and schedule overruns and their root causes (Mansfield *et al.* 1994; Assaf, Alkhalil and Alhazmi 1995; Chan and Kumaraswamy 1997; Flyvbjerg *et al.* 2003; Sambasivan and Soon 2007; Scott-Young *et al.* 2008). A few common themes emerge from these studies. First, the quality and the extent of early design and planning are key to project performance. Second, factors that influence the quality of task implementation, from experience to technical complexity, are critical contributors to overall performance. Third, late changes requested by clients often lead to many ripple effects that cost the project beyond the direct cost of those changes. Finally, the leadership and team structure and incentives moderate the impact of project-specific factors on performance. Nevertheless, different survey designs, and recall and other biases associated with retrospective studies, complicate the quantitative integration of these findings to tease out different causal mechanisms responsible for project performance and to provide quantitative decision support.

A second research stream has applied simulation modeling to understanding project dynamics and improving project management (Lyneis *et al.* 2007; Mingers *et al.* 2010). From task interdependence, to design and implementation quality, testing, and intermittent change requests, projects involve tightly interconnected factors that interact over time to determine project performance. Starting with a model that informed the arbitration of a ship-building project lawsuit (Cooper 1980), this line of modeling has grown to be one of the most successful areas of system dynamics (SD) practice (Lyneis *et al.* 2007). The rework cycle—the phenomenon of low work quality resulting in rework and change orders that extend required resources and duration—is at the core of project models and has ample empirical support (Hanna, Russell, Gotzion *et al.* 1999; Hanna and

4

Gunduz 2004; Ibbs 2005; Moselhi, Assem and El-Rayes 2005; Alnuaimi, Taha, Al Mohsin *et al.* 2010; Jarratt, Eckert, Caldwell *et al.* 2011). Moreover, from early on the modelers identified the importance of disaggregating these models to include multiple phases or task groupings (Lyneis *et al.* 2007). Formulating multi-phase project models was discussed in detail by Ford and Sterman (1998) and many applications have used different variants of this formulation in different industries (Lee, Han and Pena-Mora 2009; Park, Kim, Lee *et al.* 2011; Khoueiry, Srour and Yassine 2013). In this formulation each phase of the project is modeled with a separate rework cycle, with the knock-on effects of the quality and progress of each phase on the successive phases. Different effects could be conceived in this setup, the most prominent of which are the impact of early phase quality on later phase productivity, the effect of early quality on later quality, and the effect of later completion of tasks on the discovery of errors in earlier phases. These effects could then activate endogenous rework, schedule pressure, and morale feedback loops within different phases, leading to much variability in project performance, quality, and costs (Ford *et al.* 1998; Lyneis *et al.* 2007). However, the strengths of these feedback mechanisms have been assumed based on qualitative knowledge or single case study estimates. Rigorous and multi-project empirical estimates are lacking in the literature, partly due to limited data. Yet such estimates are critical for understanding the relative impact of different dynamic mechanisms that underlie project performance heterogeneity.

A few statistical studies have analyzed the latent impact of design error on the construction phase. Baruti and colleagues (1992) reported that design defects are responsible for 79% of total change costs, and 9.5% of total project cost. Cusack (1992)

showed that documentation errors increase project costs 10%. Hanna and colleagues (2002) found that design errors lead to 38%-50% of change orders in the projects they studied. And recently, Lopez and Love (2012) showed that the average of direct and indirect costs for design errors is about 7% of contract value. Nevertheless, estimates that clearly delineate the different causal pathways through which design errors influence the quality and productivity of construction work are lacking.

Moreover, the impact of these feedbacks is most relevant in the context of their interaction with other feedback mechanisms in the project dynamics. For example, design quality problems may lead to construction delays exacerbating burnout and morale problems, which can lead to further deterioration of construction quality and a more salient rework problem. Capturing such interactions is necessary for understanding the many instances of late and failed projects and requires a dynamic modeling framework that provides reliable estimates of the interacting factors.

In this study we develop a system dynamics model of project dynamics, empirically estimating both project-specific parameters and industry-wide inter-phase feedbacks. In light of the important roles these feedback effects play in many project models, a reliable quantitative estimate will deepen our theoretical understanding of the causal pathways in project performance, allow us to assess the relative role of these feedbacks in project performance heterogeneity, and strengthen the practical models for project planning and project dispute resolution. Moreover, our methodological approach provides a blueprint for estimating and using project-specific and industry-wide

parameters of dynamic models across other contexts, such as software, energy, infrastructure, military, and aerospace projects.

2- Data and Methods

In this study, we quantify the design-construction feedback relationships in design-bid-build (DBB) construction projects. In contrast to concurrent design and construction, DBB is a project delivery method in which design services to produce construction documents (CD Design) are performed separately from, and before, actual construction. Moreover, in the construction phase a design firm—usually the same firm that was hired for the design phase—is hired to provide inspection and design services during construction. While overlapping design and construction is common in construction projects due to their time savings (Ford and Sterman 2003), the institution owning the projects in our sample opted for DBB to reduce the risk of unanticipated iterations.

A generic dynamic model with two phases of design and construction is developed based on the SD literature. Historical data from 30 building construction projects is used to estimate and validate the model. The model is calibrated with 15 randomly selected projects and the other 15 projects are used for validation. The calibration process is used to estimate three distinct effects: 1) impact of design quality on construction quality, 2) the effect of design quality on construction productivity, and 3) the effect of construction progress on error discovery rate in design. The validation process verifies the feasibility of using simple SD models to estimate the likely distribution of project outcomes for new projects, a key step in project planning activities.

2.1- Data

A sample of thirty small-to-medium-sized projects was selected from a public university construction project archive. Data from DBB projects with the same facility type (educational building) were collected, leading to a sample of projects with \$0.5-\$60 million in budget. The dataset includes, for each project, the (initially) estimated duration (duration based on planning), initially estimated cost, actual duration, actual cost, and the cost change trajectory of the project over time (based on owner payments), all separated by the design and construction phases. The sample statistics for estimated time-to-finish (F_0), the ratio of actual to estimated time to finish (F/F_0), estimated cost (W_0), and the ratio of actual to estimated cost (W/W_0) are shown in Table 1 for design (D) and construction (C) phases of calibration and validation projects.[1] Projects in our sample show, on average, 10-50% schedule or cost overrun, depending on the phase and subsample. Coming from the same organization, industry, project type, and size, this sample allows us to control for some of the factors contributing to performance variation but unobservable in our data. However, the homogeneity of our sample also limits the generalizability of the results.

Table 1: Descriptive statistics of calibration (n=15) and validation (n=15) data

---

[1] Variable names start with the phase (C for Construction and D for Design) followed by the descriptive concept.

Sample	Variable	Unit	Mean	Std Dev	Median	Minimum	Maximum
Calibration	D_F0	Month	10.9	7.70	5.83	3.5	26.3
	D F/F0	Dmnl	1.4	0.37	1.22	1.0	2.1
	D W0	$ (M)	1.48	1.59	0.60	0.07	4.42
	D W/W0	Dmnl	1.3	0.31	1.21	0.9	2.2
	C F0	Month	32.0	18.39	31.10	9.4	65.0
	C F/F0	Dmnl	1.4	0.47	1.42	0.9	2.6
	C W0	$ (M)	18.14	20.56	6.35	0.35	55.06
	C W/W0	Dmnl	1.1	0.29	1.05	0.7	1.8
Validation	D F0	Month	12.0	10.81	10.43	3.3	47.2
	D F/F0	Dmnl	1.2	0.26	1.07	1.0	1.7
	D W0	$ (M)	1.12	1.27	0.58	0.11	4.06
	D W/W0	Dmnl	1.1	0.19	1.11	0.9	1.7
	C F0	Month	27.7	16.64	24.97	11.0	68.6
	C F/F0	Dmnl	1.5	0.39	1.43	1.0	2.1
	C W0	$ (M)	12.06	14.28	6.37	0.35	42.32
	C W/W0	Dmnl	1.1	0.07	1.07	0.9	1.2

2.2- Model development

We developed a system dynamics construction project model that was informed by the literature, first author's direct observations in the field, and a few interviews with project stakeholders. The model is developed at the level of design and construction phases. In each phase, the completion of tasks was followed by a review process (called Design Review and Inspection in the design and construction phases, respectively (Figure 1)). In each phase, the work and review activities are modeled utilizing a simple rework cycle formulation (Richardson and Pugh 1981)(Richardson G. P. and Pugh, 1981)(Richardson G. P. and Pugh, 1981)(Richardson G. P. and Pugh, 1981)(Richardson G. P. and Pugh, 1981)(Richardson G. P. and Pugh, 1981). While more complex rework cycle formulations exist (e.g., see Rahmandad and Hu 2010), the simple formulation, with three stocks for Work To Do, Accepted Work, and Undiscovered Rework, is consistent with the level of aggregation available from our data, which does not include details on individual tasks. Moreover, we allow negative values for rework to capture

scope reduction. However, scope reduction is conceptually different from rework and thus is fed into a separate stock that is kept out of the rework cycle, while the positive rework flows through the Undiscovered Rework stock and later shows up for rework. The value of the negative rework stock is subtracted from Work To Do to reflect the actual scope at any point in time.
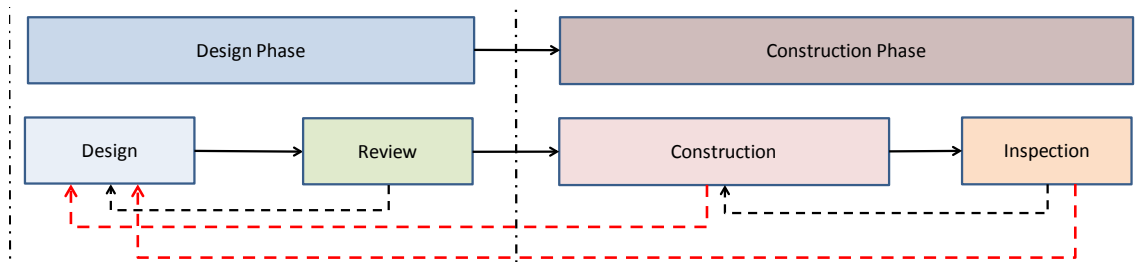

Figure 1: Construction project work flow

Figure 2 overviews the model developed in Vensim™. The model captures two phases of design and construction in two separate rework loops. The project-specific parameters of production rate (P), error rate (E), and time to detect undiscovered rework (D), are normalized by project initial values (i.e., initial work (W_0) and Duration (T_0)) to make them comparable across different projects.

In practice, the starting and finishing of each DBB project is regulated by five events: 1) Design Start, 2) Construction Document (CD) Finish, 3) Construction Start, 4) Construction Finish, and 5) Design Service (DS) Finish. Design Start, *D_Start,* is the event that initiates design. Design finishes when construction documents (CD) are approved and delivered for the bidding process. We track the time at which the design phase is perceived to be complete by the variable *DCD_Finish.* We consider the design phase as complete when no known task is left to do, even though some undiscovered rework may exist. Design CD finish triggers the start of the bidding process, during

10

which neither design nor construction activities progress, and therefore we do not include this period in our models. The next event, the construction start (*C_Start*), commences at the end of the bidding process. Construction proceeds until the construction finish, *C_Finish*, event occurs, when the construction progress passes a 99% threshold of scope. Meanwhile, some design rework, undiscovered at the design finish, will be discovered and fixed during the construction phase.

In DBB projects, usually the same architectural and engineering (A/E) designer who did the initial design is recruited to provide design services (DS) during the construction phase; therefore the initial design and later design services could be seen as the same process and are both represented within a single stock-and-flow diagram. The last event is the end of design services, *DDS_Finish*, which is triggered when the approved design work passes the 99% threshold of scope. We use our data to specify the *D_Start* and *C_Start* events for each project, while the *DCD_Finish, C_Finish,* and *DDS_Finish* are all endogenously calculated through simulation.

Inter-phase relationships between design and construction have been discussed in prior literature. Some researchers have proposed the design rework/error as the main contributor to (lack of) construction quality. Lyneis and Ford (2007) call this the Errors Build Errors effect as the quality of downstream work, e.g. Construction, is reduced by undiscovered errors in upstream work, e.g. design. They include consulting evidence (Cooper 1980; Lyneis, Cooper and Els 2001) and several academic studies as examples in which these inter-phase effects are explicitly included (Abdel-Hamid and Madnick 1991; Rodrigues and Bowers 1996; Ford *et al.* 1998; Ford, Anderson, Damron *et al.* 2004).

Others have proposed design change as the main contributor to reduced construction labor productivity (Ibbs 1997; Hanna *et al.* 1999; Hanna *et al.* 2002; Hanna *et al.* 2004; Ibbs 2005; Moselhi *et al.* 2005). Ford and Sterman (1998) developed a four-phase product development project model, in which they identified three inter-phase interactions: 1) work progress in upstream activities constrains progress in downstream activities, 2) downstream work is corrupted by upstream errors, and 3) coordination between downstream and upstream activities is required to correct errors.

In specifying the key feedbacks, we augmented the literature with qualitative data. The first author interviewed three senior project managers who had been involved in several of the projects in our sample. The field experts were individually interviewed in five separate sessions. We first provided them with a brief background of notations and basic ideas in project modeling, including the Rework Cycle. The experts were then asked to list the important design and construction interactions based on their experience and to support each with a real example. At the end of the interviews, we asked our experts to rank the first three high-impact causal links across the two phases. Consistent with the prior literature, the following inter-phase relationships were overwhelmingly selected by our panel as the most important mechanisms, worthy of further consideration.

1. Undiscovered design rework may increase construction error rate
2. Undiscovered design rework may slow down construction production rate
3. Construction progress may increase the detection rate of undiscovered design rework

We therefore model these three feedback mechanisms between design and construction phases. We capture the first knock-on mechanism, *Error Domino Effect*, in

Equation 1. The parameter S_{ED} specifies the strength of the Error Domino Effect, which we estimate as an industry-wide (i.e., project-independent) parameter. We assume the construction error rate is a function of the undiscovered rework in the design phase multiplied by a project-specific parameter representing base construction error rate, detailed in Equation 2. The undiscovered design rework (*D_UndiscoveredRework*) is normalized against initial design scope, DW_0. As noted above, we allow negative values for rework to capture scope reduction, but account for these separately in a stock for scope reduction.

Equation 1

Error Domino Effect = 1/(1+D_UndiscoveredRework / DW_0)^{SED}

Equation 2

*C_InfluencedErrorRate = Min(1,Max(-1, CE * Error domino effect))*

In the absence of data on human resources allocated to the project, in each phase a single productivity parameter is used to capture both the number of project employees and the productivity per full-time equivalent (FTE) employee. While this factor, *DP* is assumed constant in the design phase for each project (but different across different projects), the construction work rate is impacted by the undiscovered rework in the previous (i.e., design) phase, through the *Slowdown Effect* (Equation 3). This is the second knock-on effect in our model and includes an industry-wide parameter for the strength of the slowdown effect, S_{SD}, that we estimate in the calibration. Equation 4 demonstrates how we calculate the work rate using a project-specific parameter for

production (*CP* or *DP*, depending on the phase; to be estimated in calibration) and two project-specific data inputs: initial work (W_0) and estimated work duration (T_0).

$$Slowdown\ Effect = 1/(1 + D\_UndiscoveredRework / DW_0)^{SSD}$$ **Equation 3**

$$C\_WorkRate = CW_0 / CT_0 * CP * Slowdown\ Effect$$ **Equation 4**

Rework discovery is assumed to happen through a first-order draining from the stock of undiscovered rework. The time constant for this delay is set as another project-specific constant (*D*) for the construction phase. However, we assume the construction progress allows faster discovery of design problems and therefore will reduce the time constant for rework discovery in the design phase. We call this third inter-phase factor the *Reality Check Effect* (*Equation 5*). We estimate the strength of the *Reality Check Effect*, S_{RC}, in calibration. *Equation 6* shows how time to detect rework (*D*) is normalized and how Reality Check Effect influences the design rework detection rate.

$$Reality\ Check\ Effect = 1/(1+C\_Progress)^{SRC}$$ *Equation 5*

$$D\_DetectionRate = D\_UndiscoveredRework / (DD * DT_0 * Reality\ Check\ Effect)$$ *Equation 6*

Figure 2 provides an overview of the causal relationships in the model. For clarity, the switches and variables that regulate the activation of different phases are not shown. Parameters that are calibrated are highlighted in bold and larger font, with

14

exogenous variables such as initial scope and schedule in underlined italics. Full model documentation (Martinez-Moyano 2012), following minimum model documentation guidelines (Rahmandad and Sterman 2012), is available in an online appendix with the complete simulation model.
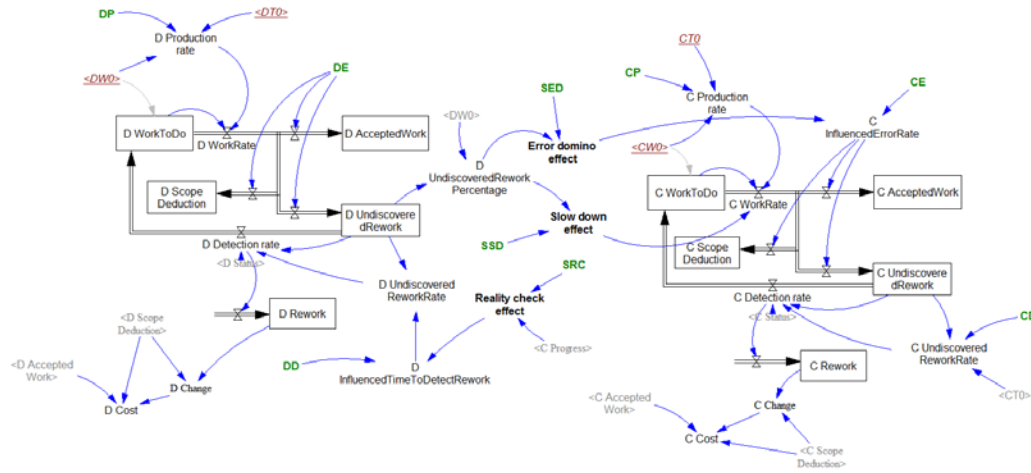

Figure 2: Overview of the model causal structure

3- Model Estimation

In the calibration step we estimate the parameters of our generic model to match the 15 randomly selected calibration projects. The estimation results are used for two purposes. First, they inform the range and distribution of project-specific parameters (i.e., error rate (E), production rate (P), and time to detect rework (D) for the two phases of design and construction). This information can then be used to form expectations on these parameter values when simulating a new project. Second, and central to the goal of this paper, we want to estimate the three inter-phase feedback effects (Error Domino, Slowdown, and Reality Check effects). These estimates are relevant both theoretically and for practical project planning purposes.

3.1- Estimation Framework

Calibration is typically conducted as a numerical optimization to estimate model parameters by minimizing the error between the model outputs and data (Oliva 2003). In our project we minimize an objective (payoff) function that is a linear combination of three sources of error: the differences between data and model in finish time, total cost, and cost curve, across both phases. Figure 3 illustrates these payoff function components.
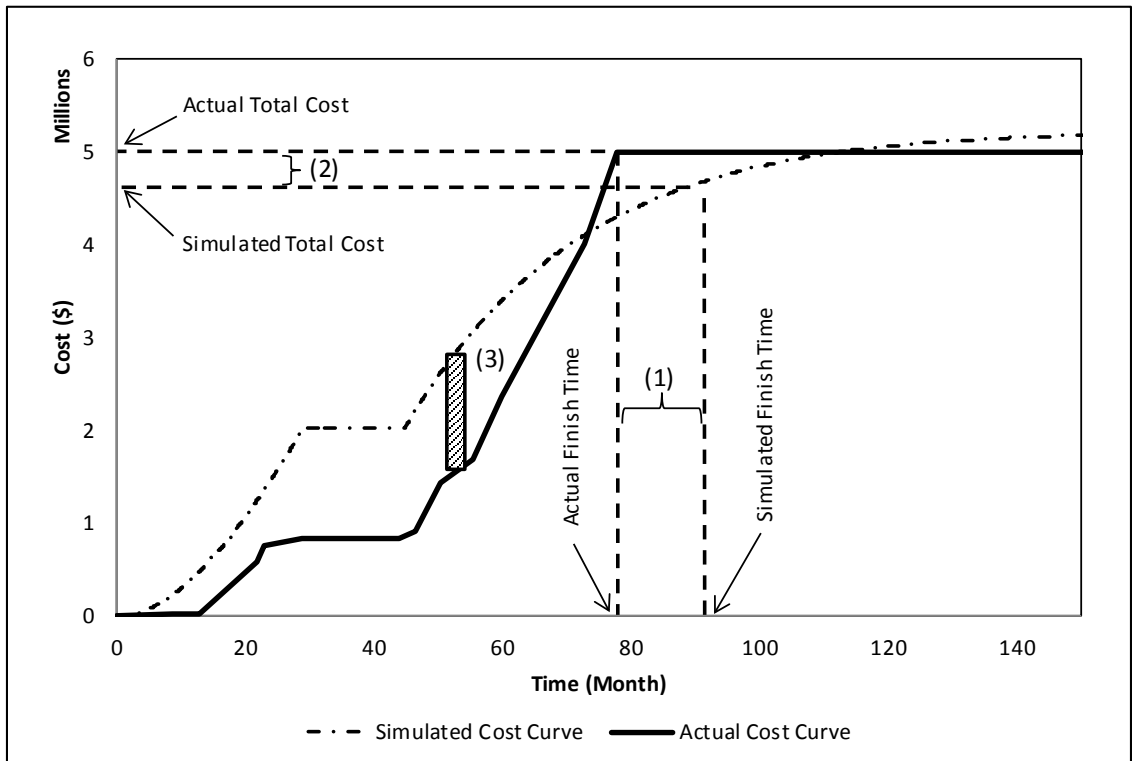

Figure 3: Calibration payoff function components

Equation 7 and Equation 8 formulate the payoff functions for design and construction, respectively. Equation 7 includes four elements for the design phase, summed over project index i: 1) the squared percentage[2] error of design construction document finish (DCD), 2) the squared percentage error of design services during

---

[2] In calculating the percentages we use the average of actual and simulated in the denominator. This avoids division by zero early in the calibration process, while keeping the payoff function robust. The alternative formulation that includes only the actual values in the denominator does not make any qualitative difference in the results but leads to more computational errors.

construction (*DDS*), 3) the squared percentage error of design total cost (*DCT*) and 4) the squared percentage error of design cost curve (*DCC(t)*). We approximate design cost curve by adding actual changes to a linear trajectory of cost distributed over project life. Equation 8 formulates the construction payoff function in the same manner, except that the construction payoff function has only one component for time, which is construction finish time (*CF*).

$$Payoff(Design)$$

$$= \sum_i \left\{ W_{DCD} \left(\frac{DCD_{sim,i} - DCD_{act,i}}{|DCD_{sim,i}| + |DCD_{act,i}|} \right)^2 \right.$$

$$+ W_{DDS} \left(\frac{DDS_{sim,i} - DDS_{act,i}}{|DDS_{sim,i}| + |DDS_{act,i}|} \right)^2 + W_{DCT} \left(\frac{DCT_{sim,i} - DCT_{act,i}}{|DCT_{sim,i}| + |DCT_{act,i}|} \right)^2$$

$$+ W_{DCC} \frac{1}{DDur_{sim,i}} \int_0^{DDur_{sim,i}} \left. \left(\frac{DCC_{sim,i}(t) - DCC_{act,i}(t)}{|DCC_{sim,i}(t)| + |DCC_{act,i}(t)|} \right)^2 dt \right\}$$

$$Payoff(Construction)$$

$$= \sum_i \left\{ W_{CF} \left(\frac{CF_{sim,i} - CF_{act,i}}{|CF_{sim,i}| + |CF_{act,i}|} \right)^2 + W_{CCT} \left(\frac{CCT_{sim,i} - CCT_{act,i}}{|CCT_{sim,i}| + |CCT_{act,i}|} \right)^2 \right.$$

$$+ W_{CCC} \frac{1}{CDur_{sim,i}} \int_0^{CDur_{sim,i}} \left. \left(\frac{CCC_{sim,i}(t) - CCC_{act,i}(t)}{|CCC_{sim,i}(t)| + |CCC_{act,i}(t)|} \right)^2 dt \right\}$$

The errors are normalized into percentages so that they can be linearly combined using weights which represent the relative importance of different components. These weights are specified subjectively based on the researchers' relative confidence in the precision of the data and the amount of information contained. For example, the cost curve is less reliable because it was reconstructed using the project invoice log, which does not perfectly match the actual completion. Therefore we reduce the weight for the cost curve and increase it for the more reliable final time and cost. Consequently, the following weights are used in the calibration results reported here: $W_{DCD} = \frac{1}{3}, W_{DDS} = \frac{1}{3}, W_{DCT} = \frac{1}{6}, W_{DCC} = \frac{1}{6}$, and $W_{CF} = \frac{1}{2}, W_{CCT} = \frac{1}{4}, W_{CCC} = \frac{1}{4}$. Finally, the design and construction payoff functions are combined with equal weights ($W_D = \frac{1}{2}, W_C = \frac{1}{2}$), to construct the total payoff to be minimized (see Equation 9). We perform some sensitivity analysis on the assumptions regarding the weights for the payoff function components and find little substantial difference within reasonable ranges for these parameters (see the section "Robustness of Calibration Results").

$$Payoff = W_D\, Payoff_{Design} + W_C\, Payoff_{Construction}$$ **Equation 9**

For calibration, 15 projects out of the 30 are randomly selected. Each project is simulated separately in the model. However, to maximize the statistical power in estimating the inter-phase feedback effects, we assume that the parameters for those effects, S_{ED}, S_{SD}, and S_{RC}, are common industry-wide and thus are the same across these 15 projects. Therefore the 15 projects are linked together through these parameters and

this requires simultaneous estimation of all projects (rather than one-by-one estimation). In general, we can classify the model parameters into two groups: 1) project-specific parameters, which are independent from one project to another, and 2) industry-wide parameters, which are common across all projects. The project-specific parameters consist of production rate (P), error rate (E), and time to detect undiscovered reworks (D) for each phase (a total of 6 parameters for each project), while the industry parameters include S_{ED}, S_{SD}, and S_{RC}. Calibration was conducted in Vensim DSS 5.8 by simultaneously estimating the project-specific and industry parameters over 15 calibration projects, leading to a total of 93 (=15*6+3) parameters to be estimated.

The large parameter space required us to perform the calibration in three phases. First, we conducted an optimization using Vensim's built-in Powell conjugate search algorithm with multiple start points in the parameter space, and using a coarse time step and convergence threshold to find a promising neighborhood for parameters. In the second phase we first fixed industry parameters, S_{ED}, S_{SD}, and S_{RC} (using values from step 1), and optimized the model, project by project, changing the project-specific parameters P, E, and D. These 15 separate calibrations provide reliable neighborhoods for all project-specific parameters. Then we fixed project-specific parameters P, E, and D and optimized the model on all projects, only allowing industry parameters S_{ED}, S_{SD}, and S_{RC} to change. These steps were repeated iteratively until parameter values converged. In phase three, we fine-tuned the results in a single optimization, starting from the point found in the previous step and allowing all the parameters to change with higher-resolution time-step and optimization settings. For more details, please see model documentation in the online appendix.

3.2- Estimation Results

Following the preceding procedure, industry parameters were estimated as S_{ED}=2.528, S_{SD}=1.262, and S_{RC}=1.188. Table 2 shows the mean, standard deviation, and correlation matrix of the project-specific calibrated parameters.

Table 2: Descriptive statistics and Correlation matrix of calibrated project-level parameters. These parameters are dimensionless.

	Mean	StdDev	D_P	D_E	D_D	C_P	C_E	C_D
D_P	0.95	0.22	1.00					
D_E	0.21	0.12	0.41	1.00				
D_D	1.32	0.81	-0.43	-0.71	1.00			
C_P	0.86	0.44	0.21	0.34	-0.47	1.00		
C_E	0.07	0.14	0.02	-0.46	0.60	-0.35	1.00	
C_D	0.43	0.50	0.21	0.05	0.09	-0.48	-0.19	1.00

Figure 4 and Figure 5 show the absolute percent error (APE) of finish time, final cost, and cost curve of design and construction, respectively, for the 15 projects. *DAPE* and *CAPE* are the weighted average errors linearly combined with the same weights used in the payoff function. The sequence of projects on the horizontal axis is based on these values sorted in descending order. Figure 6 and Figure 7 report the best and worst fits among the calibrated projects.
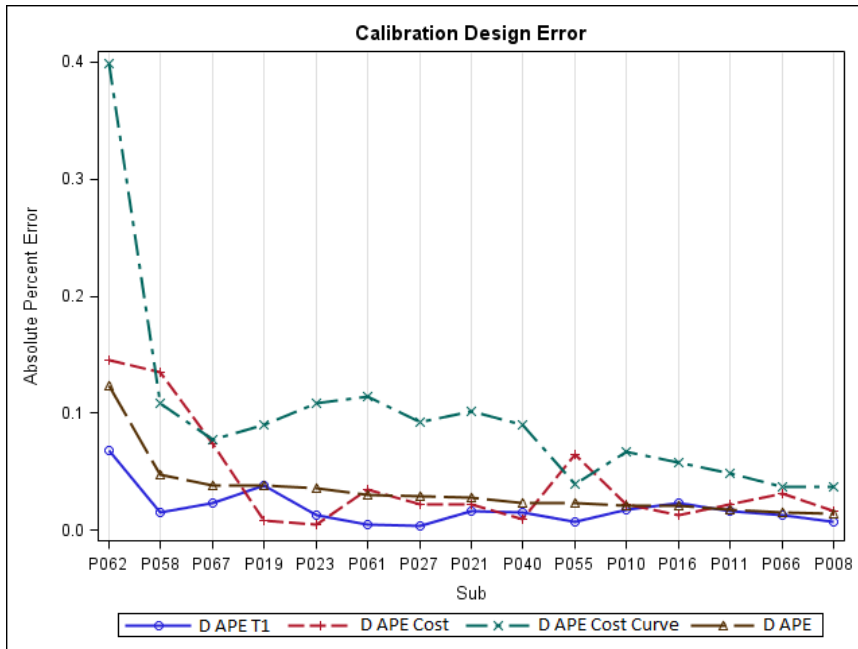
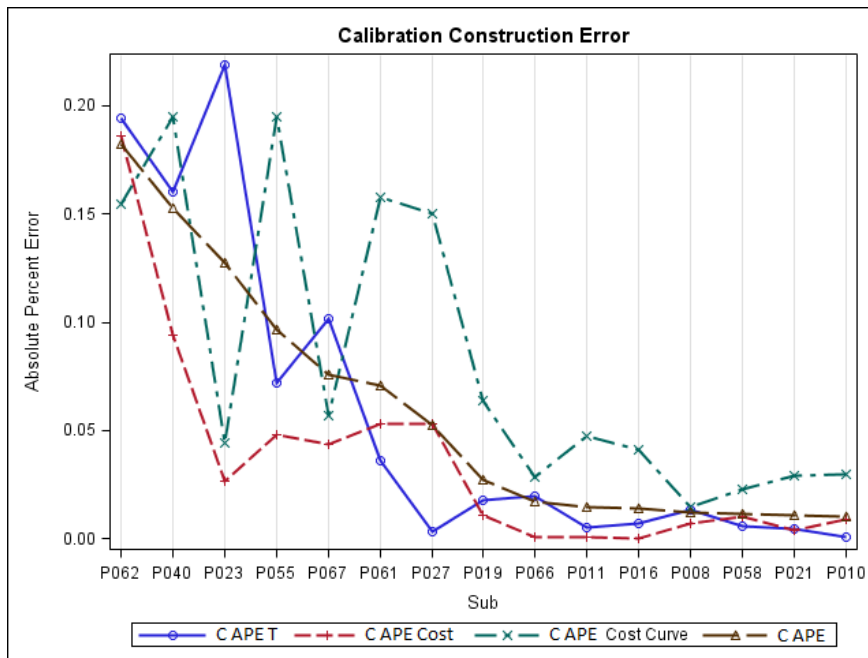**Figure 4: Design calibration error**

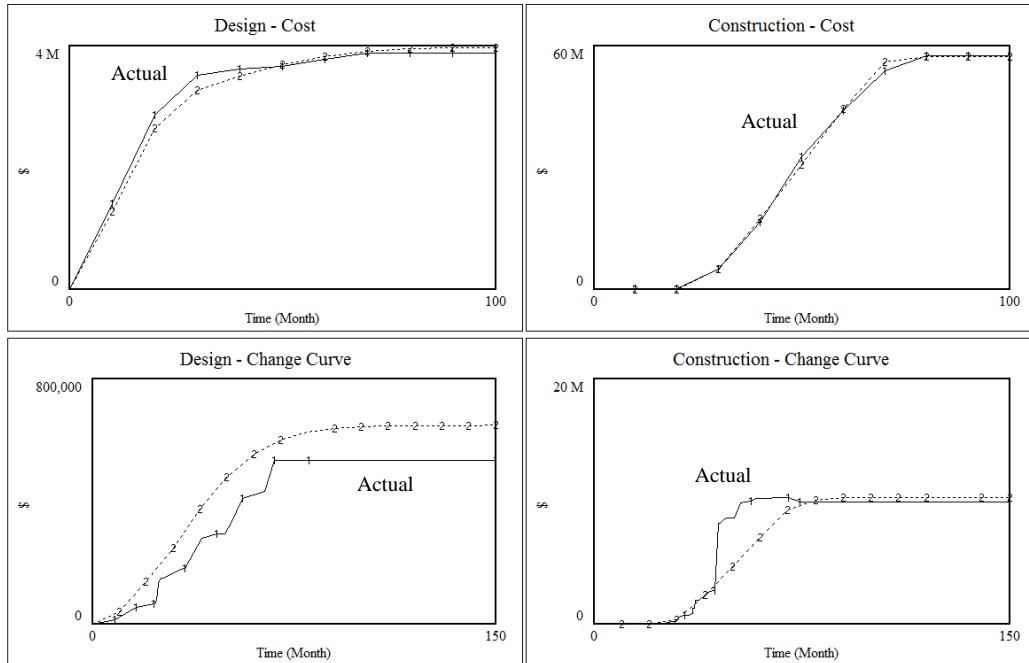**Figure 5: Construction calibration errors**

Figure 6: Simulation result of Project P011 (Best fit). Design CD Finish = 23.4 (Simulated), 23.8 (Actual). Construction Finish = 74.5 (Simulated), 74.2 (Actual)
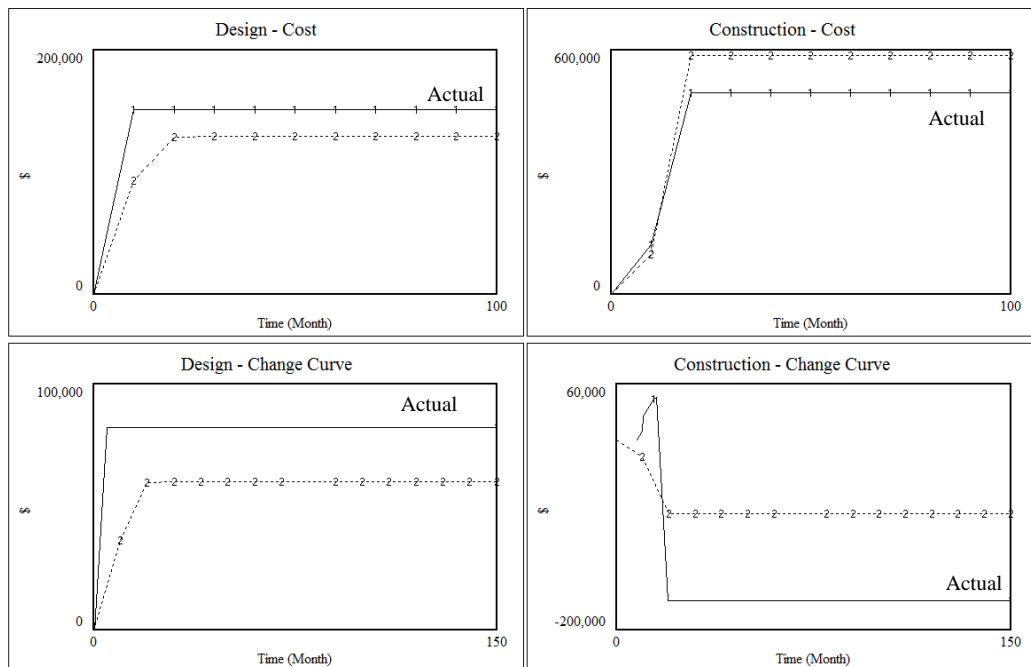

Figure 7: Simulation result of Project P062 (Worst fit). Design CD Finish = 6.6 (Simulated), 7.0(Actual). Construction Finish = 17.6 (Simulated), 19.9 (Actual)

3.3- Robustness of Estimation Results

We conduct two sets of sensitivity analyses to assess the robustness of calibration results. First, we evaluate the confidence we can have in the values reported for the feedback parameters S_{ED}, S_{SD}, and S_{RC}. Specifically, we change these parameters around their estimated value and measure the fractional change in the payoff. In the absence of a formal maximum likelihood interpretation for the payoff function, we heuristically use a 5% change in payoff as a threshold that signals incongruence between the parameters and the data.[3] The results offer the following approximate feasible ranges for the three parameters: $0.2<S_{ED}<4.2$, $0.56<S_{SD}<1.64$, and $0.91<S_{RC}<1.39$. All effects remain positive in this confidence interval. The tightest estimates belong to S_{RC} (the Reality Check effect), followed by S_{SD} (the Slowdown effect).

A second sensitivity analysis is conducted to assess the weighting functions used in defining the calibration payoff. Ten different scenarios are defined with different sets of weights listed in Table 3. The model is re-calibrated under each scenario. The impact of error weight on different scenarios is calculated by the average of the absolute percentage change of calibrated parameters. The results show no more than 6% variation in this metric across all scenarios. These findings suggest that the calibration results are robust in the reasonable range of payoff weights.

---

[3] While the complex non-parametric structure of the distributions rules out theoretical proofs, we think the 5% threshold is reasonable. For demonstration, consider a maximum likelihood-based payoff function with normally distributed errors (which, similarly to our setting, leads to normalized squared error terms in the log-likelihood function). For a sample with N effective data points (i.e., total data points minus the number of parameters), the range of a typical log-likelihood function at the best-fit position is (roughly speaking, being a chi-square distribution with N degrees of freedom) around N. In such a setting, depending on the confidence levels required, a reduction of approximately 4 units in the log-likelihood (i.e., 4/N in fractional terms) signifies reasonable confidence intervals (e.g., around 95%). With an N value well above 100 in our setting (15 projects each having 5 single data points and 2 time series of approximately 10 data points, minus the 93 parameters estimated), a fractional change in payoff of 5% provides realistic approximate bounds on the estimated parameters.

Table 3: Scenarios of error weight sensitivity analysis

Scenario	W_{D_CD}	W_{D_DS}	W_{D_CT}	W_{D_CC}	W_{C_F}	W_{C_CT}	W_{C_CC}
1	1	1	1	1	1	1	1
2	1	1	2	1	1	2	1
3	1	1	1	2	1	1	2
4	20	20	1	1	2	1	1
5	2	2	10	1	2	1	1
6	2	2	1	10	2	1	1
7	2	2	1	1	20	1	1
8	2	2	1	1	2	10	1
9	2	2	1	1	2	1	10
10	20	20	10	10	2	1	1

3-4. Inter-phase Project Feedback Effects

Figure 8 shows the magnitude of the inter-phase feedback impacts on construction error rate (diamonds in Panel a), construction production rate (squares in Panel a) and time to detect undiscovered design rework (Panel b) using the calibrated values (S_{ED}=2.528, S_{SD}=1.262 and S_{RC}=1.188). Looking at the simulations across our full sample, we note that the fraction of undiscovered design error of initial design work (x-axis in panel a) does not exceed 20%. This range confines the Error Domino and Slowdown effects to about 50% and extrapolation outside of this range is not warranted based on the current analysis. The Reality Check effect's input, construction progress, ranges in the full scale of 0 to 1 and influences rework discovery time in the design phase.
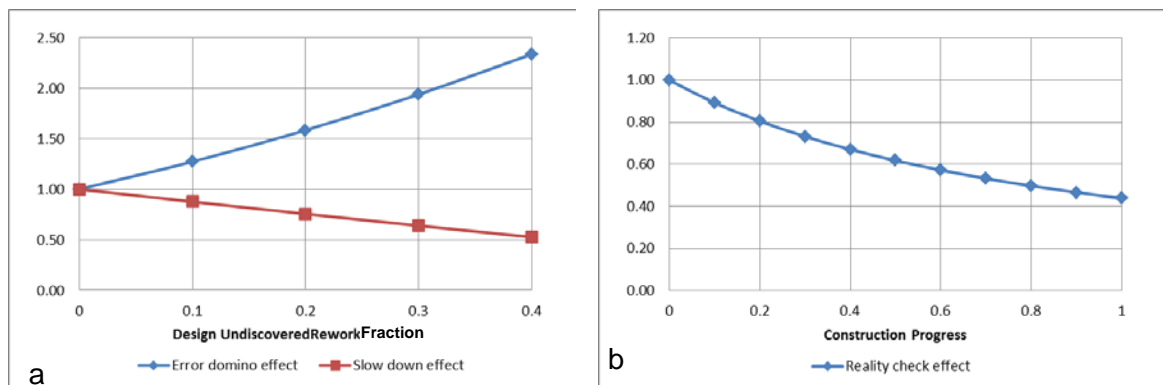
4- Predicting the Performance of New Projects

Project planning could be an important use-case for any dynamic model of projects. The parameters estimated in the previous section can be used to forecast the trajectory of a new project before it has started, potentially offering enhanced prediction and risk assessment for project planning. A simple approach is to use the averages of the 15 sets of calibrated parameters for the parameters of the prediction model. This approach, however, ignores the significant variability observed in the parameters across different projects. Ignoring the variability would give more confidence to the projections than is warranted and deprive the user of much valuable information regarding the expected distribution of potential performance outcomes. Therefore, we use a more realistic approach that assumes the six project-specific parameters are random variables with a given mean and covariance structure, which can be found from our estimated parameters. We will then generate 1000 samples with the same mean and covariance matrices for these six parameters using the variance-covariance method.

The estimated parameters are correlated, with statistics reported in Table 2, and their distribution is skewed. After some exploratory analysis we find log-normal to be a good distribution to characterize them. We therefore first apply a logarithmic transformation to the estimated project-level parameters and calculate the mean (μ) and covariance ($\sum$) of the resulting multivariate normal distribution. Next, using Equation 10, a multivariate normal sample, R, is produced that matches the joint distribution of the (transformed) calibrated parameters. Here R_{0-1} is standard normal distribution and matrix

U is the square root of the covariance matrix, $\sum$, calculated by Cholesky decomposition method (Golub and Van Loan 1996). Finally, the resulting sample is transformed back into log-normal distributions, generating R_T, to be used in the Monte-Carlo simulation experiments.

$$[R]=[\mu]+ [U]*[R_{0-1}] \qquad \text{Equation 10}$$

$$\text{Where: } [\sum]=[U]^T [U] \qquad \text{Equation 11}$$

Next, a Monte-Carlo simulation generates the distribution of model outcomes using sample R_T and a given initial plan (i.e., DW_0, CW_0, DT_0, and CT_0). Figure 9 shows the simulation result for an example project against actual outcomes. Initial scope and schedule are typically available at the beginning of any project, but are unreliable and often underestimate the actual costs and schedule significantly. These estimates are the only project-specific inputs we need in our model to generate the ensemble of possible performance projections for a new project. The project above is simulated with the 1000 sets of randomly generated parameters discussed above. Of the samples, 11% (113 out of 1000) were found infeasible as they did not result in design and construction completion in a reasonable amount of time. Comparing actual project outcomes against remaining simulations, we find design and construction times slightly higher than the predicted simulated sample median, and costs lower than the median. In all cases the actual values are within the 95% confidence levels from the simulated samples. The general fit is reasonable, given that many other relevant factors such as project type (new/renovation), location, and complexity, were not considered in this analysis.

The failure of some of the simulated projects to reach completion results from parameter combinations that lead to very high error rates and/or low productivity, and extend the project duration beyond reasonable ranges. This could reflect both inaccuracies in the parameter distributions as well as real-world mechanisms that may lead to failure of actual projects. In fact, many projects fail to reach completion in practice due to feedbacks that compromise the quality and productivity of troubled projects (Repenning 2001; Taylor and Ford 2006). However, given the small sample we use for calibration and the resulting uncertainty in the generation of joint distributions, the fraction of failed projects in this simulation is not a reliable indicator of actual projects' propensity to fail.
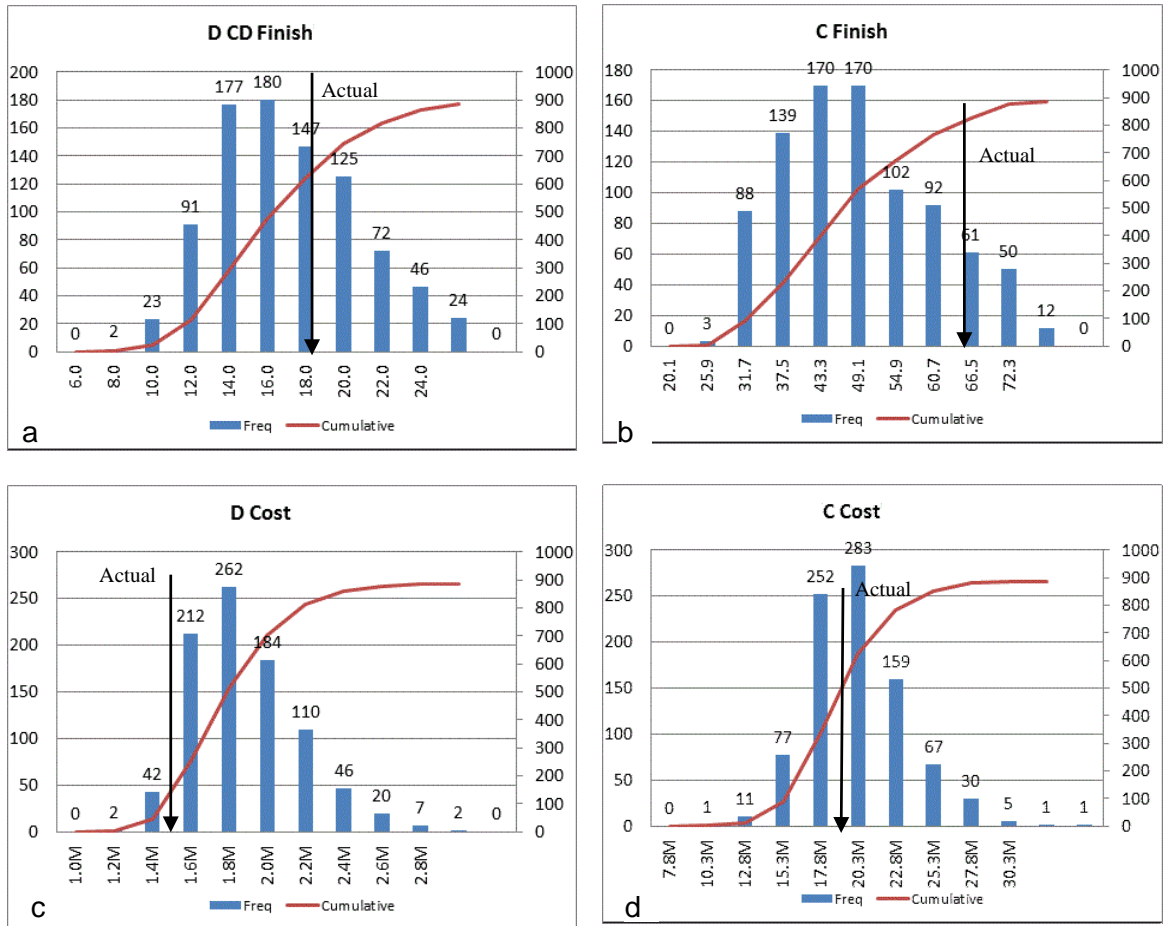
Figure 9: Comparison of actual outcomes for one validation project against Monte-Carlo simulations. a) Design finish, Mean=16.1, StDev=3.72, Actual=18.1(month) b) Construction finish, Mean=46.3.0, StDev=11.76, Actual=65.9(month) c) Design final cost, Mean=$1.8M, StDev=$0.28M, Actual=$1.5 d) Construction final cost, Mean=$18.9M, StDev=$3.15M, Actual=$18.5M

4.1- Validation: How well can new projects be predicted?

Building on the ideas above, we now formally assess the ability of the model to predict

the actual performance of new projects, given their original scope and schedule.

Specifically, we repeat the Monte-Carlo process above for the 15 validation projects,

using the 887 feasible random parameter sets. We consider four metrics, including

construction document finish time (*DF*), design cost (*DC*), construction finish time (*CF*),

and construction final cost (*CC*). The distribution of samples produced by the Monte-

Carlo simulation should be compared with the actual values for each metric and each

28

project. To simplify the presentation of results and comparisons across different projects, the simulated metrics are normalized against the actual values so that the value 1 represents the true value. These results are reported in Figure 10 in boxplot format displaying the interquartile range, mean (plus symbol), and median (solid line), and the maximum and minimum. The actual values never fall outside of the prediction envelopes and systematic biases are hard to detect by eye. To further explore the existence of such biases, we consider how well the overall distribution of predicted metrics matches the simulated distributions.
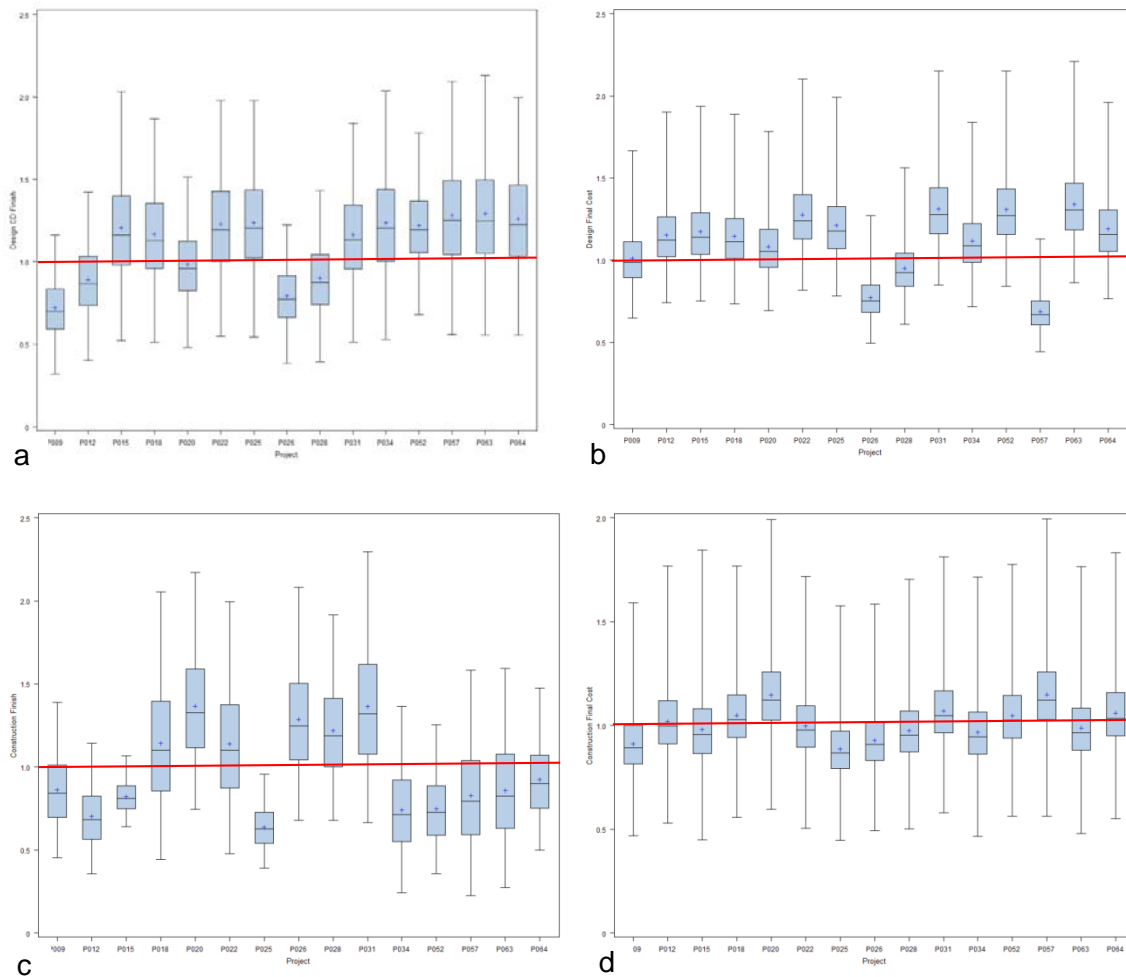

Figure 10- Mean (plus symbol), median (line), inter-quartile range, minimum and maximum for simulated metrics of each project, normalized around true value (1). a) design construction document finish (*DF*) b) design final cost (*DC*) c) construction finish (*CF*) d) construction final cost (*CC*)

The best predictive model is the one which not only gets the performance measures correct on average (i.e., has no bias in the mean across many samples), but also correctly estimates the variability expected in performance. For example, the model would have been overestimating the variance if the model's mean performance always matched the actual numbers (i.e., all boxes were set squarely on value 1), because the projected variability in outcomes was not borne out by the data. On the other hand, if most boxes were above, or below, the true value line, we would identify a bias in the model's predictions. To better assess the overall fit of the projected model metrics against the validation data, we create a variant of Q-Q plot which combines the data from all four metrics and 15 projects into a single diagnostic graph. Consider n=60 (15*4) actual metrics and their corresponding simulated distributions obtained through the Monte-Carlo results above. First, we find what percentile each data point belongs to on the corresponding simulated distribution. The resulting data set includes 60 data points with different percentile values. We sort this data set in the ascending order of percentiles and graph its values on the x-axis against the y-axis of k/(n+1) for data point k (see Figure 11). A perfect match will be on the 45-degree line, where the empirical metrics match the corresponding percentiles in simulation distributions exactly. A bias is identified if the graph is generally above or below the 45-degree line. A line much steeper than 45 degrees suggests the model is overestimating the variation in the actual metrics, i.e., it predicts many far-fetched values that never actually materialize in practice. Conversely, a line less steep than 45 degrees signals the model's overconfidence, i.e., projecting as unlikely the values that are seen regularly in practice. Finally, the goodness of fit between

a linear model and the data indicates how close the overall distribution of actual outcomes is to the predictions.
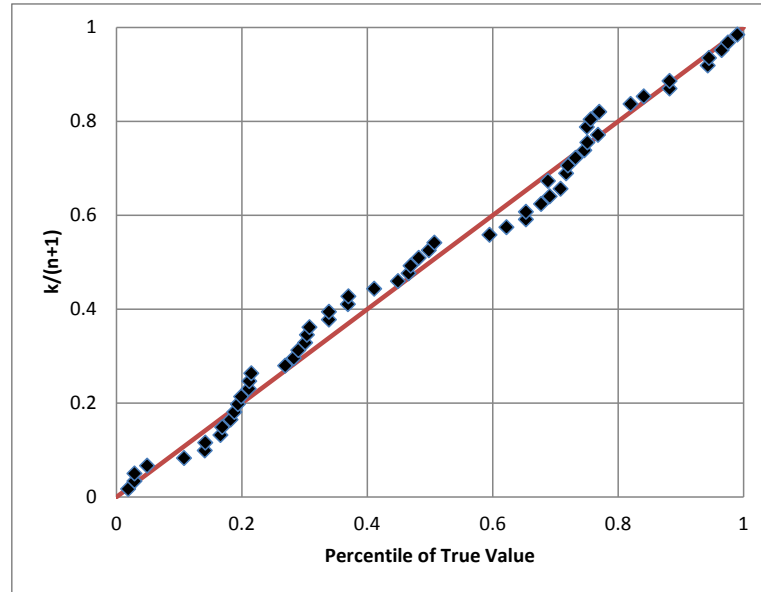

Figure 11: Q-Q plot, Percentile of true value against uniform distribution

The linear regression analysis performed on the Q-Q data series shows a very good fit between the data and the regressed line with R^2 of 0.99. Moreover, the deviations of estimated values (intercept=0.014 with standard deviation of 0.008 and slope=0.976 with standard deviation of 0.014) from the theoretical values for a perfect model (intercept=0 and slope=1) are not statistically significant at a 95% confidence level. Combined with the high quality of fit, these metrics suggest that the overall distribution of the predicted outcomes is indistinguishable from the actual distributions in the validation set. Therefore, this analysis provides further evidence in support of the viability of our predictive model: not only does our model offer a good prediction for the actual outcomes, but it also indicates the variance that can be expected in the outcomes of the projects.

Despite the reliable fit provided by our prediction method, the range of variation for the predicted metrics is notable. Essentially, the model predicts much potential variation in projects due to factors that are not endogenously captured in the current simulation, but are reflected in the uncertainty in the key project-level parameters. Therefore, absence of data on other relevant factors such as project type, location, and complexity may explain the large range of variation in predicted outcomes. If such data is available, it could be integrated into our framework as predictors of productivity, quality, and error discovery time parameters, offering a natural connection between the traditional project estimation methods and the dynamic framework we propose.

5- How important are inter-phase feedback effects?

How much of the variation in project schedule and costs can be attributed to the feedback loops we have quantified? The feedback effects we identify are only a subset of factors relevant to project heterogeneity. From intra-phase feedbacks (such as burnout, morale, and corner cutting), to other inter-phase feedbacks we did not consider, and diverse project-specific factors (e.g., size, complexity, novelty, management, organization), many other issues are relevant to understanding why projects may end up with different outcomes. Our model provides one way to tease out the effects of the feedbacks we quantify. Specifically, a good estimate of these impacts requires one to run experiments in which only the specific feedback loops are turned on and off, keeping everything else constant, so that their unique contribution can be measured. Such controlled experiments are not feasible in actual projects, but could be conducted here because we have estimated the distributions of project-level parameters that capture different contributors to variation, and therefore we can control for those factors. We

simulate five different scenarios with the ensemble of 887 feasible project-level parameter sets (see section 4), and switching on/off the three feedback mechanisms. In the base scenario we remove all the feedback effects ($SED=SSD=SRC=0$), in scenarios 1-3 we include them one at a time, and in scenario 4 we include all the feedbacks simultaneously. We then measure the change in the mean of main project outcomes (schedule and cost, for each phase, and total) as well as the change in the standard deviation of the outcomes. While the change in mean indicates how much the various feedbacks contribute to delay and costs, the latter change measure indicates how much of variability in project outcomes is due to each reinforcing loop. These results are reported in Table 4.

Table 4: The percent change of mean and standard deviation in duration and cost across different scenarios as compared to the scenario with no feedback.

Scenario	Metric	Design Duration	Const. Duration	Total Duration	Design Cost	Const. Cost	Total Cost
Scenario 2 Error Domino	Δ Mean	0.0%	2.5%	1.6%	0.0%	1.2%	1.0%
	Δ StDev	0.0%	2.7%	2.7%	0.0%	23.3%	24.9%
Scenario 3 Slowdown	Δ Mean	0.0%	8.5%	5.4%	0.0%	0.0%	0.0%
	Δ StDev	0.0%	2.5%	2.5%	0.0%	0.1%	0.1%
Scenario 4 Reality Check	Δ Mean	-19.2%	0.0%	0.0%	0.1%	0.0%	0.0%
	Δ StDev	-44.9%	0.0%	0.0%	0.0%	0.0%	0.0%
Scenario 5 All Feedbacks	Δ Mean	-17.9%	9.2%	5.8%	0.1%	0.9%	0.8%
	Δ StDev	-43.5%	4.2%	4.2%	0.0%	18.7%	20.0%

Overall, the feedbacks have a relatively small effect on the average cost (~1%) and a modest but notable impact on average duration (5.8%). They provide a more important explanatory mechanism for understanding heterogeneity, explaining 4.2% of variation in schedule and 20% of cost variation. The biggest contributor to cost variation is the Error Domino effect. In fact, the other two feedbacks have very little impact on the

cost variation. The Slowdown effect is more relevant to understanding the changes in duration of projects, and variation in duration. The impact of Reality Check feedback is rather limited and largely related to the duration of design phase. Finally, noting the higher total cost and cost heterogeneity in Scenario 2 compared to Scenario 5, we can identify an interesting interaction between Error Domino and Slowdown effects. In the presence of the Slowdown effect (Scenario 5), the lower productivity leads to longer construction duration (9.2% vs. 2.5% in Scenario 2), which in turn increases the amount of design error discovered compared to Scenario 2. Increased error discovery helps avoid some of the rework in construction, which reduces the overall construction cost in Scenario 5 compared to 2, even though the total duration of the project is higher in the presence of both feedbacks.

While these effects are not huge, they reflect the independent effect of the inter-phase feedbacks we estimate after excluding many relevant feedback effects within each phase of the project as well as non-feedback sources of variation, from project size and complexity to technology. Hence, the fact that inter-phase feedbacks may actually be explaining as much as 20% of heterogeneity in project costs is notable.

6- Discussion and Conclusions

Theoretical and Empirical Findings. We empirically estimate three design-construction feedback relationships and show that 1) undiscovered design rework increases construction error rate (Error Domino effect) 2) undiscovered design rework slows down construction production rate (Slowdown effect), and 3) construction progress increases the detection rate of undiscovered design rework (Reality Check effect). To our

knowledge this is the first empirical estimation of these feedback relationships, which we identify using empirical data from 15 construction projects. The empirical estimates validate qualitative hypotheses in this domain and suggest that the inter-phase feedback mechanisms on quality, productivity, and rework discovery time are important and of a magnitude that can make a significant impact on project dynamics. On average, we find these feedbacks may be explaining as much as 20% of variation in project costs and 6% of delays. The impact of these feedback mechanisms on a specific project's performance also depends on the base values of project quality, and to a lesser extent, productivity. For example, cutting one of these feedbacks in projects with high error rates ($CE \sim 30\%$ in our sample) leads to 5-10% overall cost variation; the effects are less pronounced for projects that start with a high quality, and thus have limited room for variation in quality due to the feedback effects.

Moreover, we find that error rates are typically higher in the design phase than construction ($DE > CE$), and take longer to be discovered ($DD > CD$). Interestingly, average productivity multipliers are higher in design ($DP > CP$), suggesting design phases that are completed on time, but with many undiscovered errors. Therefore, we expect higher emphasis on the design phase in many projects would prove fruitful. The estimated feedback effects explain a notable share of heterogeneity in project costs (~20%). We also find that much project variation originates from differences in the base value of quality and productivity. In our setting these base values aggregate diverse underlying factors, such as employee skills, project complexity, technology, management quality, and dynamic factors such as corner cutting, among others (Lyneis *et al.* 2007);

more detailed data can allow for unpacking project performance heterogeneity further using a similar simulation and calibration framework.

Given the interactions between the baseline quality and the inter-phase feedbacks, we expect these feedback processes to be most important in explaining the fate of projects in the tail of performance distribution. While many projects are completed close to initial plans, tales of projects gone wrong are common and attract much public scrutiny (Pear *et al.* 2013). Our study highlights the risk of one relevant failure mode, identified by several other researchers (Repenning 2001; Repenning, Goncalves and Black 2001; Ford and Sterman 2003; Taylor *et al.* 2006), that is rooted in multiple reinforcing mechanisms: some projects spend too little on upfront design quality, leading to problems that are harder to discover and much more costly to fix down the line. The resulting unexpected rework pushes the project further out of control, so that multiple aspects of the project suffer simultaneously: design problems hurt the relationship between design and construction and slow rework; schedule pressure may lead to burnout, corner cutting, and further quality erosion; and loss of experienced employees escalates the rework costs (Hanna *et al.* 2002; Lyneis *et al.* 2007). Moreover, the resulting threats to reputation and job security may trigger defensive routines inside project organizations, promote a culture of opaqueness, and put at risk the future projects managed by the same organization (Ford *et al.* 2003). These dynamics may be intuitive, but time and again prove central to the troubles of organizations across different industries (Repenning 2001; Repenning and Sterman 2002). In fact, if impervious to these mechanisms, following the traditional project management mindset, which seeks above all to keep the project on plan, may actually exacerbate the problems by pushing the team to work harder and focus on the

deadline, rather than finding and fixing the root causes of emerging problems (Herroelen and Leus 2004; Williams 2005).

Practical Implications. The managerial implications of these dynamics may seem straightforward: project managers need to put more emphasis on upfront design activities and ensure high quality early on. As a corollary, they should also be sensitive to signs of burnout, fear, mistrust, and communication breakdown in their teams. In fact, managing work pressure and resource loading should be pursued with a focus on avoiding such defensive organizational routines. Moreover, encouraging early discovery and revelation of problems and instituting root-cause analysis, automated testing, and other quality-inducing capabilities can help set the right tone inside the organization. Finally, given the large set of uncertainties involved, managers should be flexible in updating initial plans when the conditions on the ground call for that; and delegate more responsibility to those on the frontlines who often have a more nuanced understanding of the actual tasks, performance, and quality, and thus can solve the problems at their root.

Operationalizing this advice may be easier said than done, however: the metrics of performance built into conventional project management tools (and education) ignore many of the relevant soft variables and the feedback mechanisms we discuss (Browning 2010). Moreover, the worse-before-better tradeoffs involved in upfront quality and organizational capability investments make it harder to learn and pursue the more flexible learning-focused style of project management (Repenning *et al.* 2002; Williams 2005; Rahmandad 2008; Williams 2008), especially in large and uncertain projects with significant sociopolitical complexities (Geraldi *et al.* 2011). Incorporating into project

tracking tools metrics that indicate the state of key feedback mechanisms may provide one avenue to operationalize this advice.

Besides better quantifying the origins of fat tails in project performance distribution and the related managerial implications, our estimates are beneficial for resolving the resulting disagreements when projects go wrong. The large and notable failures in projects frequently lead to legal disputes and require model-based assessment of root causes of those failures to resolve such disputes (Cooper 1980; Ackermann *et al.* 1997; Stephens, Graham and Lyneis 2005). These modeling applications require the allocation of cost overruns to factors for which customers were responsible (e.g., change orders), contractor responsibilities (e.g., errors), and various ripple effects due to project feedback mechanisms. Since those ripple effects are relatively large, often larger than the costs directly attributable to customers or contractors (Cooper 1980), the quantitative estimates for the strength of various feedback loops are essential for a fair allocation of costs. Our estimates provide more reliable measures to quantify a subset of feedbacks central to those models.

Methodological Contributions. Two novel methodological features of this study may prove helpful for future research. First, this paper provides a method for rigorous representation of project risks through simulation. The majority of operations research models of project planning offer a deterministic schedule (Herroelen *et al.* 2004), and the ones that consider risks often assume independence among different risk factors (Taroun 2014). However, these risks are highly interdependent: on the one hand, quality, productivity, and other project-specific parameters share many common determinants,

and thus they are highly correlated. On the other hand, the feedback processes active in the evolution of a project over time couple different risks. For example, a project that falls behind on quality early on is more likely to face additional risks down the line due to corner cutting, error domino, slowdown, and other feedbacks. By estimating the covariance among project-specific parameters and explicitly modeling the feedbacks, we have provided a potential method for quantifying and tracking interdependent risks in projects. The resulting model performs very well to match calibration sample projects as well as the distributions of validation data. More realistic cost-benefit analysis, portfolio planning, and risk management can then be based on such distributions of likely outcomes. In fact, our model can be easily used for providing both baseline and risk estimates in construction projects with similar scope and complexity, and the basic approach can be replicated to inform other types of projects.

Finally, this study provides a methodological innovation to better extend system dynamics modeling to multiple cases. In fact, building models that tackle a class of problems, rather than a single case, is among the founding principles of System Dynamics (Forrester 1961). However, in practice, given the data availability and computational complexities, SD models are often calibrated using data from a single case. In a few instances where data from more than one case study is available, the model is separately estimated for those cases, treating every model parameter as distinct across different cases (e.g., Homer 1987; Sterman 1989). Such treatment foregoes the statistical power and generalizability achievable by treating the subset of parameters that transcend different cases and thus can have a case-independent value. In the current study we show how one can leverage multiple case studies to estimate both case-specific and case-

independent parameters in dynamic models, using a method that conceptually resembles fixed-effect regression. Given the general nonlinear structure of common SD models, computational costs and non-convexity of the payoff function could prove challenging in this method; nevertheless, careful design of optimization procedure—e.g., using the multi-stage method we followed here—can offer a feasible heuristic for finding reliable parameter estimates.

Limitations and Future Research. This study includes several limitations that can motivate future research. We did not explicitly consider many potentially relevant factors such as project type (new/renovation), location, and project complexity. Such factors may impact the project behavior and cost curve, and moderate the feedback effects of interest in our setting. Predictions might become more accurate, if such data were available and used in the calibration-validation process. Future studies can combine surveys of projects with the quantitative and time series data such as those we use here to estimate the root causes of variations in productivity, error rate, and rework discovery parameters and offer more explicit managerial recommendations based on those. Experiments may also be designed to elicit various effects more directly (Bendoly, Swink, & Simpson, 2014). Our data was limited to small-to-medium-sized DBB construction projects in the U.S., limiting the generalizability of our findings, especially the project-specific parameters and their covariance matrix, to other settings. In fact, the use of these parameters for prediction in project settings substantially different from ours is not warranted. Given the low to moderate structural, dynamic, and sociopolitical complexity (Geraldi *et al.* 2011) of our setting, our estimates may be rather conservative in quantifying the impact of inter-phase feedbacks on overall project performance.

Additional complexity can interact with these feedbacks, leading to more extreme cases of failure; however, the basic mechanisms should remain similar, and thus one may expect the estimated individual feedback effects to be less variable with project complexity. Finally, more detailed data about the type and scope of different tasks and resources allocated to them can enhance the quality of estimation and offer avenues for better integration of system dynamics models of projects with traditional project management software.

Acknowledgements

References:
Abdel-Hamid T, Madnick SE. 1991. *Software project dynamics: An integrated approach*. Englewood Cliffs, NJ, Prentice-Hall.
Ackermann F, Eden C, Williams T. 1997. Modeling for litigation: Mixing qualitative and quantitative approaches. *Interfaces* **27**(2): 48-65.
Alnuaimi AS, Taha RA, Al Mohsin M, Al-Harthi AS. 2010. Causes, Effects, Benefits, and Remedies of Change Orders on Public Construction Projects in Oman. *Journal of Construction Engineering and Management-Asce* **136**(5): 615-622.
Assaf SA, Alkhalil M, Alhazmi M. 1995. Causes of Delay in Large Building Construction Projects. *Journal of Management in Engineering* **11**(2): 45-50.
Bendoly E, Swink M. 2007. Moderating effects of information access on project management behavior, performance and perceptions. *Journal of Operations Management* **25**(3): 604-622.

Bendoly, E., Swink, M., & Simpson, W. P. (2014). Prioritizing and Monitoring Concurrent Project Work: Effects on Switching Behavior. Production and Operations Management, 23(5), 847–860.

Browning TR. 2010. On the alignment of the purposes and views of process models in project management. *Journal of Operations Management* **28**(4): 316-332.

Burati JL, Farrington JJ, Ledbetter WB. 1992. Causes of Quality Deviations in Design and Construction. *Journal of Construction Engineering and Management-Asce* **118**(1): 34-49.

Callahan MT, Bramble BB, Lurie PM. 1990. *Arbitration of construction disputes*. New York, Wiley Law Publications.

Chan D, Kumaraswamy M. 1997. A comparative study of causes of time overruns in Hong Kong construction projects. *International Journal of Project Management* **15**(1): 55-63.

Checkland P, Winter M. 2006. Process and content: two ways of using SSM. *Journal of the Operational Research Society* **57**(12): 1435-1441.

Cooper KG. 1980. Naval Ship Production: A Claim Settled and a Framework Built. *Interfaces* **10**(6).

Cusack D. 1992. Implementation of ISO 9000 in Construction. ISO 9000 Forum Symposium, Institution of Engineers. Gold Coast, Australia.

Drezner JA, Jarvaise JM*, et al.* 1993. An analysis of weapon system cost growth. CHARLESTON AFB SC., AIR FORCE MOBILITY CENTER.

Flyvbjerg B, Holm MKS, Buhl SL. 2003. How common and how large are cost overruns in transport infrastructure projects? *Transport Reviews* **23**(1): 71-88.

Ford DN, Anderson SD*, et al.* 2004. Managing constructibility reviews to reduce highway project durations. *Journal of Construction Engineering and Management-Asce* **130**(1): 33-42.

Ford DN, Sterman JD. 1998. Dynamic modeling of product development processes. *System Dynamics Review* **14**(1): 31-68.

Ford DN, Sterman JD. 2003. The Liar's Club: Concealing rework in concurrent development. *Concurrent Engineering-Research and Applications* **11**(3): 211-219.

Ford DN, Sterman JD. 2003. Overcoming the 90% syndrome: Iteration management in concurrent development projects. *Concurrent Engineering-Research and Applications* **11**(3): 177-186.

Forrester JW. 1961. *Industrial Dynamics*. Cambridge, The M.I.T. Press.

General Accounting Office. 2002. Transportation Infrastructure: Cost and Oversight Issues on Major Highway and Bridge Projects. Washington, DC, GAO-02-702T.

Geraldi J, Maylor H, Williams T. 2011. Now, let's make it really complex (complicated) A systematic review of the complexities of projects. *International Journal of Operations & Production Management* **31**(9-10): 966-990.

Golub GH, Van Loan CF. 1996. *Matrix computations*. Baltimore, Johns Hopkins University Press.

Gunasekaran A, Ngai EWT. 2012. The future of operations management: An outlook and analysis. *International Journal of Production Economics* **135**(2): 687-701.

Hagen M, Park S. 2013. Ambiguity Acceptance as a Function of Project Management: A New Critical Success Factor. *Project Management Journal* **44**(2): 52-66.

Hanna AS, Camlic R, Peterson PA, Nordheim EV. 2002. Quantitative definition of projects impacted by change orders. *Journal of Construction Engineering and Management-Asce* **128**(1): 57-64.

Hanna AS, Gunduz M. 2004. Impact of change orders on small labor-intensive projects. *Journal of Construction Engineering and Management-Asce* **130**(5): 726-733.

Hanna AS, Russell JS, Gotzion TW, Nordheim EV. 1999. Impact of change orders on labor efficiency for mechanical construction. *Journal of Construction Engineering and Management-Asce* **125**(3): 176-184.

Herroelen W, Leus R. 2004. Robust and reactive project scheduling: a review and classification of procedures. *International Journal of Production Research* **42**(8): 1599-1620.

Homer JB. 1987. A Diffusion Model with Application to Evolving Medical Technologies. *Technological Forecasting and Social Change* **31**(3): 197-218.

Hong P, Nahm AY, Doll WJ. 2004. The role of project target clarity in an uncertain project environment. *International Journal of Operations & Production Management* **24**(11-12): 1269-1291.

Ibbs CW. 1997. Quantitative impacts of project change: Size issues. *Journal of Construction Engineering and Management-Asce* **123**(3): 308-311.

Ibbs W. 2005. Impact of change's timing on Labor productivity. *Journal of Construction Engineering and Management-Asce* **131**(11): 1219-1223.

Jarratt TAW, Eckert CM, Caldwell NHM, Clarkson PJ. 2011. Engineering change: an overview and perspective on the literature. *Research in Engineering Design* **22**(2): 103-124.

Khoueiry Y, Srour I, Yassine A. 2013. An optimization-based model for maximizing the benefits of fast-track construction activities. *Journal of the Operational Research Society* **64**(8): 1137-1146.

Lee S, Han S, Pena-Mora F. 2009. Integrating Construction Operation and Context in Large-Scale Construction Using Hybrid Computer Simulation. *Journal of Computing in Civil Engineering* **23**(2): 75-83.

Lopez R, Love PED. 2012. Design Error Costs in Construction Projects. *Journal of Construction Engineering and Management-Asce* **138**(5): 585-593.

Lyneis JM, Cooper KG, Els SA. 2001. Strategic management of complex projects: a case study using system dynamics. *System Dynamics Review* **17**(3): 237-260.

Lyneis JM, Ford DN. 2007. System dynamics applied to project management: a survey, assessment, and directions for future research. *System Dynamics Review* **23**(2-3): 157-189.

Mansfield N, Ugwu O, Doran T. 1994. Causes of delay and cost overruns in Nigerian construction projects. *International Journal of Project Management* **12**(4): 254-260.

Martinez-Moyano IJ. 2012. Documentation for model transparency. *System Dynamics Review* **28**(2): 199-208.

Mingers J, White L. 2010. A review of the recent contribution of systems thinking to operational research and management science. *European Journal of Operational Research* **207**(3): 1147-1161.

Moløkken K, Jørgensen M. 2003. A review of surveys on software effort estimation. IEEE International Symposium on Empirical Software Engineering, Rome, Italy.

Moselhi O, Assem I, El-Rayes K. 2005. Change orders impact on labor productivity. *Journal of Construction Engineering and Management-Asce* **131**(3): 354-359.

Oliva R. 2003. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research* **151**(3): 552-568.

Park M, Kim WY, Lee HS, Han S. 2011. Supply chain management model for ready mixed concrete. *Automation in Construction* **20**(1): 44-55.

Pear R, LaFraniere S, Austen I. 2013. From the Start, Signs of Trouble at Health Portal. The New York Times. http://www.nytimes.com/2013/10/13/us/politics/from-the-start-signs-of-trouble-at-health-portal.html.

Pollack J. 2007. The changing paradigms of project management. *International Journal of Project Management* **25**(3): 266-274.

Rahmandad H. 2008. Effect of delays on complexity of organizational learning. *Management Science* **54**(7): 1297-1312.

Rahmandad H, Hu K. 2010. Modeling the rework cycle: capturing multiple defects per task. *System Dynamics Review* **26**(4): 291-315.

Rahmandad, H., & Sterman, J. D. (2012). Reporting guidelines for simulation-based research in social sciences. System Dynamics Review, 28(4), 396–411.

Repenning NP. 2001. Understanding fire fighting in new product development. *The Journal of Product Innovation Management* **18**: 285-300.

Repenning NP, Goncalves P, Black LJ. 2001. Past the tipping point: The persistence of firefighting in product development. *California Management Review* **43**(4): 44-+.

Repenning NP, Sterman JD. 2002. Capability traps and self-confirming attribution errors in the dynamics of process improvement. *Administrative Science Quarterly* **47**(2): 265-295.

Richardson GP, Pugh AL. 1981. *Introduction to system dynamics modeling with DYNAMO*. Cambridge, Mass., MIT Press.

Rodrigues A, Bowers J. 1996. System dynamics in project management: A comparative analysis with traditional methods. *System Dynamics Review* **12**(2): 121-139.

Sambasivan M, Soon Y. 2007. Causes and effects of delays in Malaysian construction industry. *International Journal of Project Management* **25**: 517-526.

Scott-Young C, Samson D. 2008. Project success and project team management: Evidence from capital projects in the process industries. *Journal of Operations Management* **26**(6): 749-766.

Stephens CA, Graham AK, Lyneis JM. 2005. System dynamics modeling in the legal arena: meeting the challenges of expert witness admissibility. *System Dynamics Review* **21**(2): 95-122.

Sterman JD. 1989. Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science* **35**(3): 321-339.

Taroun A. 2014. Towards a better modelling and assessment of construction risk: Insights from a literature review. *International Journal of Project Management* **32**(1): 101-115.

Tavares LV. 2002. A review of the contribution of Operational Research to Project Management. *European Journal of Operational Research* **136**(1): 1-18.

Taylor T, Ford DN. 2006. Tipping point failure and robustness in single development projects. *System Dynamics Review* **22**(1): 51-71.

Williams T. 2005. Assessing and moving on from the dominant project management discourse in the light of project overruns. *Ieee Transactions on Engineering Management* **52**(4): 497-508.

Williams T. 2008. How do organizations learn lessons from projects - And do they? *Ieee Transactions on Engineering Management* **55**(2): 248-266.

Winter M. 2006. Problem structuring in project management: an application of soft systems methodology (SSM). *Journal of the Operational Research Society* **57**(7): 802-812.