A Theory of Cost Overruns

Neha Jain*and Ram Singh†

October 3, 2012

(An incomplete draft.)

Abstract

Cost overruns in public procurement are a global phenomenon. The literature on the subject has proposed several models to explain cost overruns. However, the existing models are 'static.' These models do not predict how cost overruns will behave overtime in terms of their frequency and magnitude. Moreover, the literature has not addressed several important issues like how the cost overruns will vary with the size of the project, and across the types of the projects, etc. In this paper, we present a unified model of cost overruns. The model predicts that cost overruns: decline over time; are relatively high for procurement involving construction projects, compared to procurement of finished products, such as machinery etc; within construction projects, more complex projects will experience higher cost overruns compared to less complex ones. Moreover, in contrast to the existing literature on incomplete contracts, we show that an increase in probability of renegotiation can increase the asking price (bid) by the bidder contractors. Predictions emanating from our model are tested with the help of a large dataset on infrastructure projects in India.

*Department of Economics, Delhi School of Economics, University of Delhi, Delhi, INDIA 110007. Email: nehajain@econdse.org.

†Department of Economics, Delhi School of Economics, University of Delhi, Delhi, INDIA 110007. Email: ramsingh@econdse.org.

1 Introduction

Cost overruns in public procurement contracts are observed worldwide. However, the literature on cost overruns, both theoretical and empirical, is scant. The existing theoretical models on cost overruns in public procurement attribute cost overruns to two factors: a) imperfect designing techniques and b) the ex-post negotiations between the government and the contractor.[1] However, these models are 'static'. Therefore, they do not say anything about how cost overruns will behave over time. These models are also silent on how cost overruns will vary across the good to be procured; the types of the projects and the sectors concerned; and whether they are influenced by other factors - such as time overruns, size of the project, etc - highlighted by the empirical findings on cost overruns.[2]

Moreover, the theoretical models predict that ex-post changes in contract terms imply greater profit for the contractor and therefore an increase in probability of contract renegotiation lowers the bids (asking price) by the bidders, since bidder bid more aggressively expecting benefits from hold up during renegotiations. However, in an empirical study Bajari, Houghton and Tadelis (2011) show just the opposite, i.e., the bids increase with the probability of contract renegotiation.

We present a model that predicts the following about cost overruns in public procurement projects.

- Cost overruns decline over time, both in terms of frequency and magnitude.
- more complex projects experience higher cost overruns compared to less complex ones.
- the cost overruns are relatively high for procurement involving construction projects, compared to procurement of finished products, such as machinery etc.
- the bids increase with the probability of contract renegotiation.

We corroborate these predictions (the first three) with the help of a large dataset on infrastructure projects in India.

[1]See Laffont and Tirole (1993), Bajari and Tadelis (2001), Ganuza (1997,2007), Bajari and Tadelis (2001), Chen and Smith(2001), Arvan and Leite (1990), and Gaspar and Leite (1989).

[2](Chen and Smith (2001); Odeck (2004), and Singh (2010)

2 Model

Procurement of public goods such as infrastructure starts with the planning and designing of the project. The planning and designing is characterized by several activities. The first is the description of the *scope* of the project. Scope of a project specifies the 'output' to be delivered. For example, for an expressway project, the scope generally specifies the length, the number of traffic lanes, the number and location of cross-section, by passes, under-passes, over-passes, cross-section, toll-plazas, major, medium and small bridges, service roads, etc. Scope for renovation may specify the extent of the repair to be done.

Once the scope of the project has been fixed, the next task is the description of works/tasks (popularly known as work-items) required to build the project facility.[3] Let

$[0, \overline{W}]$ denote the set of all possible works needed to be performed for the given scope.

Plausibly, $0 < \overline{W}$ and $\overline{W}$ varies across project scopes. Even for the given scope of the project, the initial design may or may not specify all of project works. Let,

W be the number of works covered by the initial design,

where $W \leq \overline{W}$. WLOG assume that initially the scope specifies the first W works; it leaves out the remaining, $(W, ..., \overline{W}]$ works. The number of initially specified works, i.e. W, depends on several factors, e.g., complexity of the project, the efforts put in project specification as well as on the experience of the planners with designing. That is, $W = W(\tau, l, \dot{d})$, where

$\dot{d}$ denotes the effort put in to determine the work-items required for the project,

l denotes the experience of designers with project planning; and

τ denotes the technical complexity of the project.

[3] A typical road project requires many works to be done; such as, construction of embankment, construction of subgrade, building of earthen and concrete shoulders, fixing of drainage spouts, laying of boulder apron, among many others. The table lists a total of 78 major and 26 minor activities for a bridge work.

Assume that $W(\tau, l, 0) = 0$, and $W(\tau, l, \infty) = \overline{W}$. That is, specification of a project work requires positive effort, and with 'sufficiently' high effort all of project works can be specified in the initial design itself. Plausibly, W is an increasing function of $\dot{d}$ and l, such that; $\frac{\partial W(\tau,l,\dot{d})}{\partial d} > 0$ and $\frac{\partial W(\tau,l,\dot{d})}{\partial l} > 0$. Besides, $\frac{\partial^2 W(\tau,l,\dot{d})}{\partial d^2} < 0$, i.e., while the initial works are relatively easy to specify, specification task becomes increasing more difficult as the index of works increases. Similarly, $\frac{\partial^2 W(\tau,l,\dot{d})}{\partial l^2} < 0$, i.e., marginal gains from the experience decrease with its length. But, $\frac{\partial W(d,l)}{\partial \tau} < 0$. Besides, $\frac{\partial^2 W(\tau,l,\dot{d})}{\partial \tau^2} \leq 0$, Finally, $\frac{\partial^2 W(\tau,l,\dot{d})}{\partial d \partial l} < 0$, i.e, higher effort is a substitute for longer experience, and vice-versa.

Next, consider a function $D = D(\tau, l, \dot{d})$ such that:

$$\begin{aligned} D(\tau, l, \dot{d}) &: \mathfrak{R}^3_+ \mapsto [0, 1]; \\ D(\tau, l, \dot{d}) &= \frac{W(\tau, l, \dot{d})}{\overline{W}} \end{aligned}$$

For the given scope of the project, the function $D(\tau, l, \dot{d})$ can be taken as a measure of the completeness of project design. The incompleteness of design can now be defined as $1 - D = 1 - \frac{W}{\overline{W}} = \frac{\overline{W} - W}{\overline{W}}$. Clearly, $D(\tau, l, 0) = 0$ and $D(\tau, l, \infty) = 1$ hold. In view of the above, the following hold: $\frac{\partial D(\tau,l,\dot{d})}{\partial d} > 0$, $\frac{\partial D(\tau,l,\dot{d})}{\partial l} > 0$, $\frac{\partial^2 D(\tau,l,\dot{d})}{\partial d^2} < 0$, $\frac{\partial^2 D(\tau,l,\dot{d})}{\partial l^2} < 0$, and $\frac{\partial^2 D(\tau,l,\dot{d})}{\partial d \partial l} < 0$.

The last designing activity is the estimation of quantities of work-items and their per-unit costs are estimated. Therefore, the project designing requires multidimensional effort on the part of designers.

The project designing is followed by tending, selecting and the signing of the procurement contract. After receiving the estimates of quantities and costs, the department publishes the project details - such as, reports describing the project scope, project works, estimates of quantities and per-unit costs - and invites the bids, say at $t = 1$. From these reports, the bidders can infer the levels of τ, l and d, values of various functions such as D, W, σ, etc. Assume that the successful bidder is chosen and the procurement contract is awarded at $t = 1$ itself. The construction starts at $t = 2$. Let

d denotes the effort in estimation of quantities of project-works and their per-unit cost.[4] Efforts d and $\dot{d}$ are put in at the beginning of project designing, say at $t = 0$. These and other efforts modeled in the paper are measured by their respective costs. Let,

q_w^a [resp. q_w^e] denote the *actual* [resp. estimated] quantity of the w^{th} work-item/activity, and

c_w^a [resp. c_w^e] denote the *actual* [resp. estimated] *per-unit* cost of the w^{th} work-item/activity.

Clearly, the actual cost of the work w^{th} is equal to

$$c_w^a \times q_w^a.$$

So, the actual costs of first W project works, i.e., of the project works covered by the initial design is given by

$$\int_0^W [c_w^a \times q_w^a] dw = \mathbf{c^a}.\mathbf{q^a},$$

where $\mathbf{q^a} = (q_w^a)$ is the vector of actual quantities and $\mathbf{c^a} = (c_w^a)$ is the vector of actual per-unit costs of first W works.

The actual costs of project works invariably turn out to be different from their estimated values. This can happen either because the actual quantities or per-unit costs or both turn out to be different from their estimated values. Below we model the relationship between the actual and the estimated costs.

The vector of actual quantities $\mathbf{q^a}$ depends on the state of nature that unfolds at the project site during the construction. For example, the type of optimum mixture of the concrete and bitumen required, the kind of foundations needed for flyovers, etc., depend on the quality of soil at the project site. Suppose, for each work w, the set of possible values for q_w^a is $(0, \overline{q}_w)$; where $w \in (0, \overline{W}]$. Similarly, for each w, let the set of possible values of c_w^a be $(0, \overline{c}_w)$.

Following the literature and to keep analysis simple, assume that the quantities and cost related contingencies get realized at the beginning of

[4] In principle, we can distinguish between the efforts in estimation of quantities and per-unit costs. However, we to keep the notational low, we do not do so.

construction, i.e., at $t = 2$. In particular, at $t = 2$ q^a_w as well as c^a_w become known for each $w \in (0, \overline{W}]$; at $t = 0$, there is uncertainty about both. However, at $t = 0$ the designers can put in effort d, as mentioned above, to estimate the values/quantities of yet unknown vectors $\mathbf{q^a} = (q^a_w)$ and $\mathbf{c^a} = (c^a_w)$, $w \in (0, \overline{W}]$.

As a result of effort d, the designers get publicly observable signals of quantity-relevant and cost-relevant states of nature. Alternatively put, the effort d produces signals/estimates of quantities, and their respective per-unit costs. Let,

q^e_w and c^e_w denote the signals of q^a_w and c^a_w, respectively.

The relations between q^e_w and q^a_w is stochastic. Specifically, q^e_w is a noisy signal of q^a_w, for each $w \in (0, \overline{W}]$. The informativeness of q^e_w about q^a_w depends on the level of d. Formally, let $q^e_w = q^a_w + \epsilon_w(q_w)$, or denoting $\epsilon_w(q_w)$ by ϵ_{q_w},

$$q^e_w = q^a_w + \epsilon_{q_w}.$$

We assume that when effort $d > 0$, q^e_w is a partially correlated signal of q^a_w.[5] In principle, the support of ϵ_{q_w} can vary across the possible values of q^a_w. So, let the support of ϵ_{q_w} be $[\underline{\xi}(q^a_w), \overline{\xi}(q^a_w)]$, where $-\infty < \underline{\xi}(q^a_w) < 0 < \overline{\xi}(q^a_w) < \infty$.[6] For each ϵ_{q_w}, $w \in (0, W]$, let

$F_{q_w}(\epsilon_{q_w}|\tau, l, d)$ be the distribution function for ϵ_{q_w};
$f_{q_w}(\epsilon_{q_w}|\tau, l, d)$ be the density functions for ϵ_{q_w}, and
$\sigma^2_{\epsilon_{q_w}}$ denote the variance of ϵ_{q_w}.[7]

Note that $\sigma_{\epsilon_{q_w}}$ is a function of d as well as l, along with the technical complexity of the project. Besides, for given values of d and l, $\sigma_{\epsilon_{q_w}}$ can, in principle, vary across the possible values of q^a_i, i.e., $\sigma_{\epsilon_{q_w}} = \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)$,

[5]When $d = 0$, assume that ϵ_{q_w} is uniformly distributed over $(-q^a_w, \overline{q}_i - q^a_w)$. That is, when no effort is put in estimation of quantities, q^e_w is not correlated with q^a_w. Recall that $q^a_w \in (0, \overline{q}_w)$.

[6]To keep the information structure simple and symmetric across various values of q^a_w, one can focus on the scenario when $q^a_w \in (0, \frac{2}{3}\overline{q})$, and $[-\underline{\xi}(q^a_w), \overline{\xi}(q^a_w)] = [-q^a_w, q^a_w]$. Note that when $q^a_w \in (0, \frac{2}{3}\overline{q})$, $q^e_i \leq \frac{3}{2}q^a_w < \overline{q}_i$, i.e., $q^e_i < \overline{q}_i$ will hold for all $w \in (0, W]$.

[7]These functions will also depend on τ.

for $i = 1, ..., W$. To keep matters simple, assume that when $d > 0$: $f_{q_w}(\epsilon_{q_w}|\tau, l, d)$ is centered around 0, i.e., $E(\epsilon_{q_w}) = 0$; and

$$F_{q_w}(\epsilon_{q_w}|\tau, l, \infty) = \begin{cases} 0 & \text{if } \epsilon_{q_w} \in (-\underline{\xi}(q^a_w), 0); \\ 1 & \text{if } \epsilon_{q_w} \in [0, \overline{\xi}(q^a_w)). \end{cases}$$

That is, q^e_i is an unbiased indicator of q^a_i, and as effort level approaches infinity, q^e_w approaches/coincides with q^a_w. Moreover,

$$\frac{\partial \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)}{\partial d} < 0 \text{ and } \frac{\partial \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)}{\partial l} < 0,$$

i.e., variance of ϵ_{q_w} is a decreasing function of d as well as l. For instance, for all $d > 0$, one can think of ϵ_{q_w} as 'normally' distributed, say, over $[\underline{\xi}(q^a_w), \overline{\xi}(q^a_w)] = [-\xi(q^a_w), \xi(q^a_w)]$, where $0 < \xi(q^a_w) < \infty$. For example, let $F_{q_w}(\epsilon_{q_w}|\tau, l, d)$ be differentiable at every $\epsilon_{q_w} \in (-\xi(q^a_w), 0)$ and every $\epsilon_{q_w} \in (0, \xi(q^a_w))$, such that:

$$(\forall l)(\forall \tau)\left(\forall \epsilon_{q_w} \in (-\underline{\xi}(q^a_w), 0)\right)\left[\frac{\partial F_{q_w}(\epsilon_{q_w}|\tau, l, d)}{\partial d} < 0\right],$$
$$(\forall l)(\forall \tau)\left(\forall \epsilon_{q_w} \in (0, \overline{\xi}(q^a_w))\right)\left[\frac{\partial F_{q_w}(\epsilon_{q_w}|\tau, l, d)}{\partial d} > 0\right],$$
$$(\forall d)(\forall \tau)\left(\forall \epsilon_{q_w} \in (-\underline{\xi}(q^a_w), 0)\right)\left[\frac{\partial F_{q_w}(\epsilon_{q_w}|\tau, l, d)}{\partial l} < 0\right],$$
$$(\forall d)(\forall \tau)\left(\forall \epsilon_{q_w} \in (0, \overline{\xi}(q^a_w))\right)\left[\frac{\partial F_{q_w}(\epsilon_{q_w}|\tau, l, d)}{\partial l} > 0\right],$$
$$(\forall d)(\forall l)\left(\forall \epsilon_{q_w} \in (-\underline{\xi}(q^a_w), 0)\right)\left[\frac{\partial F_{q_w}(\epsilon_{q_w}|\tau, l, d)}{\partial \tau} > 0\right],$$
$$(\forall d)(\forall l)\left(\forall \epsilon_{q_w} \in (0, \overline{\xi}(q^a_w))\right)\left[\frac{\partial F_{q_w}(\epsilon_{q_w}|\tau, l, d)}{\partial \tau} < 0\right].$$

Then, $\frac{\partial \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)}{\partial d} < 0$, $\frac{\partial \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)}{\partial l} < 0$ and $\frac{\partial \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)}{\partial \tau} > 0$ will hold. That is, for any given level of l and τ, if $d'_1 > d$ then $\sigma_{\epsilon_{q_w}}(q^a_i, l, d'_1) < \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)$ will hold. Similarly, for any given level of d and τ, if $l' > l$, then $\sigma_{\epsilon_{q_w}}(q^a_i, l', d) < \sigma_{\epsilon_{q_w}}(q^a_i, \tau, l, d)$ will hold. However, *ceteris paribus*, the variance increases with complexity. Clearly, the distribution $F_{q_w}(\epsilon_i|l, d'_1)$ [resp. $F_{q_w}(\epsilon_i|l', d)$] second order stochastically dominates the distribution $F_{q_w}(\epsilon_i|\tau, l, d)$ [resp. $F_{q_w}(\epsilon_i|\tau, l, d)$].

The per-unit cost of a work item depends on the cost of inputs (material, labour, capital, etc) as well as various efforts that are put by the entity responsible for construction, generally a contractor. We model this effort. Let,

a denote the construction costs reducing effort.

The literature suggests that organization of project works is crucial for the construction costs. a can be interpreted as the expenditure or efforts in organization of works, searching and securing supply of inputs, manpower, etc. By improving the efficiency of work, a reduces construction costs. Specifically, for work w, a reduces per-unit construction costs by $\kappa_w^1(a)$. We assume that

$$\frac{\partial \kappa_w^1(a)}{\partial a} > 0, \ \& \ \frac{\partial^2 \kappa_w^1(a)}{\partial a^2} < 0.$$

That is, construction cost decrease with the management-effort by the contractor. We assume that the effort a is a *scope-specific* investment by the contractor, in the sense described later. Therefore, for a work-item w, the actual per-unit cost is given by

$$c_w^a(a) = \kappa_w^0 - \kappa_w^1(a),$$

where κ_w^0 can be interpreted as the per-unit cost of meeting contractually specified standard and specifications of work w, in the absence of the efforts a by the construction contractor.

Returning to the expected per-unit cost, c_w^e, we assume that for any given level of a, c_w^e is partially correlated and noisy signal of c_w^a. Let ϵ_{c_w} denote the error term for the stochastic relationship between c_w^e and c_w^a. So, for any given level of a, we have

$$c_w^e = c_w^a + \epsilon_{c_w}$$

As above, assume that when $d = 0$, ϵ_{c_w} is uniformly distributed over $(c_w^a, \bar{c}_w - c_w^a)$. However, when effort $d > 0$, let the support of ϵ_{c_w} be $[\underline{\xi}(c_w^a), \overline{\xi}(c_w^a)]$, where $-\infty < \underline{\xi}(c_w^a) < 0 < \overline{\xi}(c_w^a) < \infty$. Let, $F_{c_w}(\epsilon_{c_w}|\tau, l, d)$, $f_{c_w}(\epsilon_{c_w}|\tau, l, d)$ and $\sigma_{\epsilon_{c_w}}(c_w^a, \tau, l, d)$, respectively, be the distribution, the density and the variance functions of ϵ_{c_w}. Assume that function $F_{c_w}(\epsilon_{c_w}|\tau, l, d)$, $f_{c_w}(\epsilon_{c_w}|\tau, l, d)$ and $\sigma_{\epsilon_{c_w}}(c_w^a, d)$ satisfy all the corresponding properties imposed above on $F_{q_w}(\epsilon_{q_w}|\tau, l, d)$, $f_{q_w}(\epsilon_{q_w}|\tau, l, d)$

and $\sigma_{\epsilon_{q_w}}(q_i^a, \tau, l, d)$, respectively.

Now, for any given level of a, the cost of first W works (ignoring the cost of efforts) is given by

$$\int_0^W [c_w^a \times q_w^a]\, dw + \int_0^W \{[c_w^a \times \epsilon_{q_w}] + [\epsilon_{c_w} \times q_w^a] + [\epsilon_{c_w} \times \epsilon_{q_w}]\}\, dw, i.e.,$$

by

$$\int_0^W [c_w^a \times q_w^a]\, dw + \epsilon, \tag{1}$$

where $\epsilon = \int_0^W \{[c_w^a \times \epsilon_{q_w}] + [\epsilon_{c_w} \times q_w^a] + [\epsilon_{c_w} \times \epsilon_{q_w}]\} dw$ is a random variable. Note that the distribution of ϵ depends on l, d, $\mathbf{c^a}$, $\mathbf{q^a}$ as well as on the distributions of ϵ_{c_w} and ϵ_{q_w}.[8] However, in view of the above it follows that, for all l and d,

$$E(\epsilon) = 0. \tag{2}$$

i.e., for any given level of a, the initially estimated costs are an unbiased indicator of the actual project costs. Let $F(\epsilon|\tau, l, d)$, $f(\epsilon|\tau, l, d)$ and $\sigma_\epsilon^2(\mathbf{q^a}, \mathbf{c^a}, \tau, l, d)$, respectively, denote the distribution function, the density and the variance of ϵ. In view of the above assumptions on $F_{q_w}(\epsilon_{q_w}|\tau, l, d)$, $f_{q_w}(\epsilon_{q_w}|\tau, l, d)$, $F_{c_w}(\epsilon_{c_w}|\tau, l, d)$, $f_{c_w}(\epsilon_{c_w}|\tau, l, d)$, $\sigma_{\epsilon_{q_w}}(q_i^a, \tau, l, d)$ and $\sigma_{\epsilon_{c_w}}(c_i^a, l, d)$, it is easy to see that

$$F(\epsilon|l, \infty) = \begin{cases} 0 & \text{if } \epsilon < 0; \\ 1 & \text{if } \epsilon \geq 0. \end{cases}$$

i.e., for any given level of a, as effort level approaches infinity the estimated costs $C^e_{(0,W]}$ approach/coincide with the actual costs $C^a_{(0,W]}$. In view of the above, we get the following proposition. That is, the variance of the error-term decreases as the effort d or the experience l increases, or both. However, *ceteris paribus*, the variance increases with complexity.

Proposition 1 $(\forall\tau)(\forall l)\left[\frac{\partial\sigma_\epsilon(\mathbf{q^a},\mathbf{c^a},\tau,l,d)}{\partial d} < 0\right]$, $(\forall\tau)(\forall d)\left[\frac{\partial\sigma_\epsilon(\mathbf{q^a},\mathbf{c^a},\tau,l,d)}{\partial l} < 0\right]$, *and*
$(\forall l)(\forall d)\left[\frac{\partial\sigma_\epsilon(\mathbf{q^a},\mathbf{c^a},\tau,l,d)}{\partial \tau} > 0\right]$.

[8]Note that the support of ϵ need not be symmetric around 0, even when both ϵ_{c_w} and ϵ_{q_w} have symmetric supports. For example, let there be $i = 1, ..., W$ works. Suppose $\epsilon_{q_w} \in [-\frac{1}{2}q_w^a, \frac{1}{2}q_w^a]$ and $\epsilon_{c_w} \in [-\frac{1}{2}c_w^a, \frac{1}{2}c_w^a]$. In that case, it is easy to check that the support of ϵ is $[-\frac{3}{4}\sum_{i=1}^n [c_w^a \times q_w^a], \frac{5}{4}\sum_{i=1}^n [c_w^a \times q_w^a]]$. For more on distributions of the product of two random or more random variables see Goodman (1962) and Lomnicki (1966).

Proof: It can be verified that

$$\begin{aligned}\sigma_\epsilon^2(\mathbf{q^a}, \mathbf{c^a}, \tau, l, d) &= \int_0^W [[(c_w^a)^2 \times \sigma_{q_w}^2(q_w^a, \tau, l, d)] + [(q_w^a)^2 \times \sigma_{c_w}^2(c_w^a, \tau, l, d)] \\ &+ [\sigma_{c_w}^2(c_w^a, \tau, l, d) \times \sigma_{q_w}^2(q_w^a, \tau, l, d)]]dw. \end{aligned} \tag{3}$$

Now, in view of our assumptions on σ_{q_w}, σ_{c_w}, etc., it is easy to see that

$$(\forall l)\left[\frac{\partial\sigma_\epsilon(\mathbf{q^a}, \mathbf{c^a}, \tau, l, d)}{\partial d} < 0\right], \ \& \ (\forall d)\left[\frac{\partial\sigma_\epsilon(\mathbf{q^a}, \mathbf{c^a}, \tau, l, d)}{\partial l} < 0\right].$$

□

3 Actual versus Estimated Costs

3.1 Construction Costs

Recall, the initial cost estimates are for the works in the set $(0, W]$. The actual cost of these works is observed at $t = 2$. As mentioned earlier, at the beginning of the construction phase the remaining $(W, \overline{W}]$ works are also incorporated in the design and completed during construction phase. For any given level of a, the actual costs of all works is given by

$$\int_0^W [c_w^a \times q_w^a]dw + \int_W^{\overline{W}} [c_w^a \times q_w^a]dw, i.e., \tag{4}$$

$$\int_0^{\overline{W}} [(\kappa_w^0 - \kappa_w^1(a)) \times q_w^a]dw. \tag{5}$$

The following proposition shows that expected cost overruns are positive. But cost overruns decrease as completeness of the design increases. Formally,

Proposition 2 *For any given level of* a*,* $E\left[\frac{C^a_{(0,\overline{W}]}}{C^e_{(0,W]}}\right] \geq 1$, $\frac{\partial E\left[\frac{C^a_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial D} < 0$, *and*
$\lim_{D\to 1} E\left[\frac{C^a_{[1,\bar{W}]}}{C^e_{[1,W]}}\right] = 1.$

Proof: From (4), for any given level of a, the actual costs of all works can be written as

$$\frac{C^a_{(0,\overline{W}]}}{C^e_{(0,W]}} = \frac{C^a_{(0,\overline{W}]}}{C^e_{(0,W]}} + \frac{C^a_{(W,\overline{W}]}}{C^e_{(0,W]}} \tag{6}$$

In view of (2)

$$E\left[\frac{C^a_{(0,\overline{W}]}}{C^e_{(0,W]}}\right] = 1 + \frac{C^a_{(W,\overline{W}]}}{C^a_{(0,W]}} \tag{7}$$

Therefore,

$$E\left[\frac{C^a_{[1,\overline{W}]}}{C^e_{[1,W]}}\right] \geq 1. \tag{8}$$

Now the rest of the claims follow from the fact that as D increases the second term in the RHS of (7) declines, since as D increases W approaches $\overline{W}$.

$\square$

The above proposition further gives us the following result.

Proposition 3 *For a given project:*

$$\frac{\partial E\left[\frac{C^a_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial d} < 0; \quad \frac{\partial E\left[\frac{C^a_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial l} < 0; \ \& \ \frac{\partial E\left[\frac{C^a_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial \tau} > 0.$$

3.2 Cost overruns for the government

Under item rate contracts, the bidders submit vectors of per-unit asking price for each work. Let $\mathbf{b} = (b_w)$ denote the vector of bids quoted by the successful bidder/contractor; where b_w is the asking price for per-unit of work item w. The contractor is paid for the actually performed quantities of work-items, according to a bid rates. So, the actual payment on account of work w will be $b_w.q^a_w)$ denote the vector of contractually agreed compensation rate. These rates are for per-unit of the relevant work item. However, since the tender documents invite bids for first W works only; the contract price for the remaining work-items $(W, \bar{W}]$ will have to be negotiated during construction. Let these prices be denoted by $(b_{w'})$ the vector of these prices, $w' \in (W, \bar{W}]$.

So, when there is no change in the scope, for given a the contractor's final payoff will be

$$\int_{w\in[0,W]} (b_w \times q^a_w)dw + \int_{w\in(W,\overline{W}]} (b_{w'} \times q^a_{w'})dw - C^a_{[0,\overline{W}]}, \tag{9}$$

where $C^a_{[0,\overline{W}]} = C^0 - \kappa(a) + a$.

Let

$$P = \int_{w\in[0,W]} (b_w \times q^a_w) dw$$

and

$$P' = \int_{w'\in(W,\overline{W}]} (b_{w'} \times q^a_{w'}) dw'.$$

So, the total final payment made by the government and received by the contractor under the UR contract are $P^{UR} + P'^{UR}$. That is, the total actual project cost for the government, C^G, is

$$C^G = P^{UR} + P'^{UR}$$

Assuming competitive bidding and rational expectations about the work quantities and the renegotiation process over remaining works items, the initial bid vector **b** will be such that it gives

$$C^G_{[0,\overline{W}]} = P + P' = C^a_{[0,\overline{W}]} \tag{10}$$

and

$$\mathbf{b}.\mathbf{q}^e = \mathbf{b}.\mathbf{q}^a = P \tag{11}$$

The following propositions follow immediately from Propositions 2 and 3, in view of (10).

Proposition 4 *For any given level of a,* $E\left[\frac{C^G_{(0,\overline{W}]}}{C^e_{(0,W]}}\right] \geq 1$, $\frac{\partial E\left[\frac{C^G_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial D} < 0$, *and*

$\lim_{D\to 1} E\left[\frac{C^G_{[1,\bar{W}]}}{C^e_{[1,W]}}\right] = 1$.

Proposition 5 *For a given project:*

$$\frac{\partial E\left[\frac{C^G_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial d} < 0; \quad \frac{\partial E\left[\frac{C^G_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial l} < 0; \;\& \; \frac{\partial E\left[\frac{C^G_{[1,\bar{W}]}}{C^e_{[1,W]}}\right]}{\partial \tau} > 0.$$

3.3 Measures and Proxies

We have data on the following variables/proxies

- $C^e_{[1,W]}$, i.e., estimated project cost (INITIALCOST)
- $C^G_{[1,\bar{W}]}$, i.e., final total project cost to the govt

- τ: We measure τ by the IMPLEMENTATIONPHASE, or the IMPLPHASE for short. It is the estimated/planned time for completion of the project. Presumably, the project planners will increase the IMPLPHASE in proportion to the complexity of the project, i.e., τ.

- l; we measure this as the time difference between the date of first project in the data-set and the date of contract award for the project at hand. By the time later projects were awarded, the designer had acquired greater experience. So, these projects have higher l (TIMELAPSE), and vice-versa.

- D: Project type- whether construction or not

- TIME-OVERRUN; the difference between the actual project completion time and the estimate project completion time

- For other variables and details see Table 1

Hypothesis 1 *Ceteris paribus, the cost overruns,* $E[\frac{C^G_{[1,\bar{W}]}}{C^e_{[1,W]}}]$,

1. *increase with TIME-OVERRUN;*
2. *increase with IMPLPHASE;*
3. *decrease with TIMELAPSE*
4. *will be higher for construction projects*

4 Change in Scope

Whether a change is scope will be required during the construction phase or not depends on the state of nature that unfolds during construction phase. The initial scope is specified assuming a particular scope-relevant state of nature. During construction if the state of nature is actually what was expected to be, then no changes in the scope is needed. In contrast, if the realized state of nature turns out to be different, then the initially specified scope has to be modified. When a change is scope is needed, the cost of changes is estimated and add to the initial cost estimates.

For instance, a road project originally could be designed to simply resurface the existing stretch without any changes in the under-surface. However, the actual site conditions may necessitate strengthening of the under- surface and shoulders. Alternatively, during construction phase the government may discover that some relevant features are missing from the original scope. For example, for a highway project government engineers may discover the need for more of flyovers or under-passes. Similarly, for a railways project the government may find that they have missed out on some safety measures in the initial design. At times, demand from local public can add to the list work-items, thereby necessitating a change in scope.[9]

The likelihood of a change in scope depends on the level of effort put in by the project designers to find out the scope-relevant state of nature and specify the scope accordingly. The project planners can reduce the likelihood of changes in the scope by putting in higher effort, $\ddot{d}$, in this regard. Formally, let

$\pi(.)$ be the probability that no change in the scope is needed.

So, the probability that change in scope will be needed during construction phase is $1 - \pi(.)$. Intuitively, $\pi(.)$ will depend on the effort $\ddot{d}$ experience of the project designer and the technical complexity of the project, i.e., $\pi(.) = \pi(\tau, l, \ddot{d})$ where

A change in scope of the project affects construction costs in two ways. One, it affects the quantities of the work items, and therefore the final costs. Two, if affects the gains from the organizational effort a. Since the works and resources are organized for the initial design, it seems plausible to argue that a is more effective in decreasing the construction if there is no change in design compared to the scenario in which the design has to be changed. Formally, effort a is a *design-specific* investment by the contractor. Let, $\hat{\kappa}^1(a)$ denote the reduction in the construction costs of work w on account of effort a, if there is change in the scope.

If there is a change in scope, for any given choice of a by the contractor, the total construction costs $\hat{C}^a_{(0,\overline{W}]}$ will be

$$\hat{C}^a = \int_0^{\overline{W}} [(\kappa^0_w - \hat{\kappa}^1_w(a)) \times \hat{q}^a_w] dw + a, \tag{12}$$

[9]Empirical studies suggest that a changes in scope, generally, leads to increases in the quantities of the existing work-items as well as bring in new tasks under the scope of the project. See for example Bajari *et al* (2011).

where $\hat{q}^a_w$ denotes the actual quantities of wth work -item after the change in scope.

Now, for given π, the expected costs are given by

$$\begin{aligned}\pi C^a + (1-\pi)\hat{C}^a &= \pi[C^0 - \kappa(a)] + (1-\pi)[\hat{C}^0 - \hat{\kappa}(a)] \\ &= [\pi C^0 + (1-\pi)\hat{C}^0] - [\pi\kappa(a) + (1-\pi)\hat{\kappa}(a)] \quad (13)\end{aligned}$$

where $C^0 = \int_0^{\overline{W}}[\kappa^0_w \times q^a_w]dw$, $\kappa(a) = \int_0^{\overline{W}}[\kappa^1_w \times q^a_w]dw$, etc

We assume the effort a to be sufficiently design specific in that the following hold:

$$\frac{\partial\hat{\kappa}(a)}{\partial a} > 0,\ \frac{\partial^2\hat{\kappa}(a)}{\partial a^2} < 0,\ \&\ \frac{\partial\hat{\kappa}(a)}{\partial a} < \frac{\partial\kappa(a)}{\partial a}.$$

So,

$$\frac{\partial[\pi\kappa(a) + (1-\pi)\hat{\kappa}(a)]}{\partial\pi} > 0$$

Proposition 6 $\left[E[\hat{C}^0_{[1,W]}] = C^0_{[1,W]}\right] \Rightarrow \frac{\partial E\left[\frac{\pi C^a_{[1,W]} + (1-\pi)\hat{C}^a_{[1,W]}}{C^e_{[1,W]}}\right]}{\partial\pi} < 0.$

That is, if the expected input costs of the alternative design is al least equal to that of the initial design, in expected terms, the ratio of the actual construction cost to the estimated costs will decline as π increases.

An argument similar to the one in the last section implies

Proposition 7 $\left[E[\hat{C}^0_{[1,W]}] = C^0_{[1,W]}\right] \Rightarrow \frac{\partial E\left[\frac{\pi C^G_{[1,W]} + (1-\pi)\hat{C}^G_{[1,W]}}{C^e_{[1,W]}}\right]}{\partial\pi} < 0.$

For the government also, in expected terms, the ratio of the actual construction cost to the estimated costs declines as π increases.

Also, note that if the vector of compensation rate in the event of design changes, $\hat{\mathbf{b}}$, is such that

$$\hat{\mathbf{b}}.\mathbf{q}^a = \hat{C}^0$$

it can be shown that

$$\frac{\partial P}{\partial\pi} < 0 \quad (14)$$

that the bid-rates (asking price) will increase with the probability of design change, as in shown in Bajari, Houghton and Tadelis (2011)

5 Project Planning, Data and Definitions

5.1 Data Description

5.2 Definitions:

For each project we can define percentage time overrun, TO, as the ratio of the actual time over the planned project completion time. Similarly, we define 'cost overrun', CO, as the ratio of the actual cost over the initially projected (i.e., expected) cost of the projects. The initially expected cost is called the initial project cost. These are cost estimates for project works and generally are arrived at using current input prices. The actual costs become known only at the time of completion at the end of phase two. A related term used in the paper is the 'implementation phase'. It is defined as the duration in which a project is planned to be completed. The 'implementation phase' is measured as the time duration from the date of award of procurement contract to the expected date of completion/execution of the project as per the contract.

The previous section, offers several testable predictions. However, to test them we need measures of project complexity, i.e., τ, as well as experience, i.e., l. Is there a general measure of complexity available? The project size seems to be a reasonable measure of complexity. Presumably the complexity increases with project size. Since, compared to smaller ones, bigger projects involve more works. The designing and coordination problems naturally increase with the number and magnitude of works, in turn, increasing the complexity. If so, our question boils down to determining the measures of project size. The data provides two measures of project size. The first is the initially estimated project cost. It seems to be a good measure of project size, its complexity, and hence of the contractual incompleteness. Following the terminology in Singh (2010), we will call the estimated project cost to be simply the INITIAL-COST.[10] The second measure is the implementation phase; the duration in which a project is initially planned to be completed. We will term

[10]The initially expected project cost, rather than the actual cost, is a better indicator of the size and incompleteness of the contract. Due to cost overrun, the final cost can be large even for small projects. The same argument applies to the implementation phase.

this measure as the IMPLEMENTATIONPHASE, or the IMPLPHASE for short. Plausibly, *ceteris paribus*, projects involving larger number of works are more complex than those requiring a smaller number. Similarly, projects with more new and complicated works are more complex than those for standard works. Obviously, as the number of works and intricacy increase, it will take longer to complete the project. Presumably, the project planners will increase the implementation phase in proportion to its complexity. In other words, the IMPLPHASE is proportional to the complexity of the project, τ.

As far as experience with project designing, l, is concerned, we measure it in term of number of months that have elapsed since the start of the first project in the sector or dataset under consideration. We call the duration as TIMELAPSE. We will denote its square by TIMELAPSESQ or TIMELAPSE^2. *Ceteris paribus*, the contractual incompleteness is expected to decrease with TIMELAPSE.

6 The Empirical Framework and Results

In view of the various definitions introduced above, our model offers the following testable hypotheses:

Hypothesis 2 *Ceteris paribus, the cost overruns and their probability will*

1. *increase with IMPLPHASE;*
2. *decrease with TIMELAPSE.*

Apart from project complexity and experience, there are other factors too that have implications for cost overruns. Delay or time or overrun in project implementation is one such factor. Arguably, any delay in implementation will cause cost overrun for the project. This can happen simply on account of inflation itself. If there are delays, inputs will become more expensive and, in turn, will cause an increase in the project cost. Moreover, certain overhead costs have to be met as long as the project remains incomplete. Delays should increase these costs also. Also, a long delay may cause depreciation of project assets, necessitating expenses on repairs or replacements. At the same time, it is pertinent to keep in mind that contract renegotiation is a time consuming and generally contested process. This means it is expected to cause not only cost overruns but also delay. If so, project complexity and experience affect the time overruns too. This suggests a simultaneity between cost and time overruns. However, as is shown in Singh (2010), while there is simultaneity between the two, the causation runs from delays to cost overruns and not the other way around.

The model and the above discussion suggests the following regression model:

$$\begin{aligned} TO &= \alpha_0 + \alpha_1 TIMELAPSE_t + \alpha_2 TIMELAPSE_t^2 + \alpha_3 INITIALCOST_t \\ &+ \alpha_4 IMPLPHASE_t + \alpha_6 PSGDP_t + \epsilon_{1t} \end{aligned}$$

$$\begin{aligned} CO &= \beta_0 + \beta_1 TIMELAPSE_t + \beta_2 TIMELAPSE_t^2 + \beta_3 INITIALCOST_t \\ &+ \beta_4 IMPLPHASE_t + \beta_6 PSGDP_t \\ &+ \beta_7 TO + \epsilon_{2t} \end{aligned}$$

7 Results

Preliminary results corroborate the predictions, but we are still working on results.

References

Table1: We have data on

S. No.	ASPECT/VARIABLE	DESCRIPTION	DATA SOURCE
1	DATE OF PROJECT START	It is the date of approval of the project	MOSPI reports and the NHAI.
2	INITIAL/EXPECTED DATE OF COMMISSIONING	It is the initially planned (i.e, expected) date of completion of the project	MOSPI reports and the NHAI.
3	ACTUAL DATE OF COMMISSIONING	It is the actual date of completion of the project	MOSPI reports and the NHAI.
4	TIMEOVERRUN	The time difference (in months) between the actual and the initially planned date of completion; Time difference b/w (3) and (2), above.	OUR CALCULATIONS based on the data collected from MOSPI reports and the NHAI. .
5	IMPLEMENTATION PHASE (IMPLPHASE) *****	The duration in which a project is planned to be completed, i.e., the duration between the date of approval of the project and its *expected* date of completion.	OUR CALCULATIONS based on the data collected from MOSPI reports and the NHAI.
6	PCTIMEOVERRUN (PCTO) *****	The ratio of the time overrun and the implementation phase for the project (multiplied by one hundred).	OUR CALCULATIONS based on the data collected from MOSPI reports and the NHAI.
7	INITIAL/EXPECTED PROJECT COST (INITIALCOST)	The initially projected (i.e., expected) cost of the project.	MOSPI reports and the NHAI.
8	ACTUAL PROJECT COST	The actual cost at the time of completion of the project.	MOSPI reports and the NHAI.
9	COST OVERRUN	The difference between the actual cost and the initially projected (i.e., expected) cost of the project.	OUR CALCULATIONS based on the data collected from MOSPI reports and the NHAI.
10	PCCOSTOVERRUN (PCCO)	The ratio of the cost overrun and the initially anticipated cost of the project (multiplied by one hundred).	OUR CALCULATIONS based on the data collected from MOSPI reports and the NHAI.
11	TIMELAPSE	It is the time (in months) that has lapsed since the date of approval of the *first* project in the relevant dataset. For all sectors projects it is the time that has lapsed since May 1974. For the set of railways projects it is the same, i.e., May 1974. For the NHAI dataset on projects it is August 1995,.	OUR CALCULATIONS based on the data collected from MOSPI reports and the NHAI.
12	SECTOR	The infrastructure sector to which the project belongs.	MOSPI reports
13	STATE	The state in which the project is located.	MOSPI reports and the NHAI and publications of

			the Ministry relevant for the Sector

* Definition for NHAI dataset is somewhat different and has been explained in the text

Table 2: Summary Statistics: All Sectors

		% Cost Overrun				% Time Overrun			
Sector	Number Of Projects	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Atomic Energy	12	15.05	113.12	-84.89	265.12	301.02	570.48	-3.13	2033.33
Civil Aviation	51	-2.07	38.97	-80.32	109.18	67.20	56.01	-12.20	289.29
Coal	102	-11.42	91.72	-99.73	466.23	30.42	69.70	-93.33	359.57
Fertilizers	16	-12.57	28.92	-67.75	50.13	26.53	41.80	-18.18	109.30
Finance	1	132.91	.	132.91	132.91	302.78	.	302.78	302.78
Health and family Welfare	2	302.30	92.96	236.56	368.03	268.04	208.63	120.51	415.56
I & B	7	14.00	62.97	-34.60	134.64	206.98	140.57	101.67	491.43
Mines	5	-33.16	20.65	-62.78	-9.88	42.44	36.23	-2.78	98.11
Petrochemicals	3	-12.22	25.92	-28.40	17.68	74.43	3.05	70.91	76.19
Petroleum	125	-15.82	29.12	-80.87	106.77	38.52	50.31	-41.67	242.86
Power	108	51.09	271.36	-61.83	2603.96	33.55	54.89	-50.00	202.08
Railways	130	94.06	178.33	-65.49	1287.98	118.05	141.13	-2.17	1100.00
Road Transport and Highways	169	14.50	61.09	-93.86	416.72	46.48	54.66	-28.26	317.39
Shipping and Ports	61	-1.35	84.35	-90.37	574.38	118.64	276.79	-7.14	2150.00
Steel	44	-15.41	47.32	-91.85	235.88	50.49	60.08	-25.00	305.56
Telecommunication	74	-33.82	56.22	-98.40	279.46	248.82	253.98	-18.18	1200.00
Urban Development	24	12.31	50.27	-48.81	144.00	66.44	44.58	3.60	166.67
Total	934	15.06	131.26	-99.73	2603.96	79.46	152.98	-93.33	2150.00

Table 3: Category-wise distribution of projects (all sectors)

Sectors/States	Number of projects
Road, Railways, and Urban-development	316
Civil Aviation, Shipping and Ports and Power	221
Inter-state; Spanning across multiple states	91
Punjab, Haryana, Delhi, Gujarat, Maharashtra,	252
A.P, Tamil Nadu, Karnataka, Kerala	222
North-East and J&K	64

TABLE : *Regression Results: ALL SECTORS*

Variables	*Model 1*		*Model 2*	
	PCGETIMEOVRRN (% Time Overrun)	*PCGECOSTOVRRN (% Cost Overrun)*	*PCGETIMEOVRRN (% Time Overrun)*	*PCGECOSTOVRRN (% Cost Overrun)*
PCGETIMEOVRRN		*0.0854 [0.0224] (0.000)*		*0.0949 [0.0220] (0.000)*
TIMELAPSE	*-2.8993 [0.3714] (0.000)*	*-2.7328 [0.3662] (0.000)*	*-2.2846 [0.3037] (0.000)*	*-2.3500 [0.3782] (0.000)*
TIMELAPSE Sq	*0.0039 [0.0006] (0.000)*	*0.0043 [0.0006] (0.000)*	*0.0029 [0.0005] (0.000)*	*0.0037 [0.0006] (0.000)*
INITIAL COST	*-0.0016 [0.0053] (0.758)*	*0.0144 [0.0033] (0.000)*		
IMPLPHASE	*-1.8848 [0.1565] (0.000)*	*0.1430 [0.1117] (0.201)*	*-1.7513 [0.1449] (0.000)*	*0.2170 [0.1058] (0.041)*
DRRU	*52.4719 [5.3119] (0.000)*	*40.0284 [3.4134] (0.000)*	*51.1584 [5.1512] (0.000)*	*37.4532 [3.3739] (0.000)*
DCSPP	*23.1145 [4.7073] (0.000)*	*20.1239 [3.3279] (0.000)*	*21.7332 [4.6854] (0.000)*	*17.5595 [3.3167] (0.000)*
DTA	*155.6228 [17.3884] (0.000)*	*-29.5410 [6.9564] (0.000)*	*159.2271 [17.5965] (0.000)*	*-33.5075 [6.6734] (0.000)*
DSTATES	*-9.5303 [6.1470] (0.121)*	*3.9355 [4.9466] (0.427)*	*-12.2252 [5.6048] (0.029)*	*4.7312 [4.6512] (0.309)*
DMRICH	*-2.6099 [5.5619] (0.639)*	*-0.4604 [3.2476] (0.887)*	*-2.9584 [5.2901] (0.576)*	*1.6575 [3.2676] (0.612)*
DRICH	*-4.8560 [5.0631] (0.338)*	*-4.7192 [3.0916] (0.127)*	*-6.2061 [4.9688] (0.212)*	*-4.2426 [3.0811] (0.169)*
DNE	*-2.8802 [7.4362] (0.699)*	*14.1414 [6.3857] (0.027)*	*3.1680 [11.1237] (0.776)*	*15.5355 [7.9015] (0.050)*
CONSTANT	*615.0301 [56.1885] (0.000)*	*383.9315 [55.8855] (0.000)*	*514.6568 [45.2008] (0.000)*	*324.6392 [58.0830] (0.000)*
Observations	*797*	*797*	*793*	*793*
R-squared	*0.4856*	*0.4521*	*0.4698*	*0.4059*

** White's heteroscedastic consistent estimates. Robust standard error in square parentheses. P-value in round parentheses.*

TABLE : *ALL SECTORS Quantile Regression*

Variables	*Model 1*		*Model 2*	
	PCGETIMEOVRRN (% Time Overrun)	*PCGECOSTOVRRN (% Cost Overrun)*	*PCGETIMEOVRRN (% Time Overrun)*	*PCGECOSTOVRRN (% Cost Overrun)*
PCGETIMEOVRRN		*0.0238 [0.0135] (0.078)*		*0.0210 [0.0108] (0.051)*
TIMELAPSE	*-1.6575 [0.1667] (0.000)*	*-3.3612 [0.1750] (0.000)*	*-1.6125 [0.1631] (0.000)*	*-3.3535 [0.1395] (0.000)*
TIMELAPSE2	*0.0019 [0.0003] (0.000)*	*0.0053 [0.0003] (0.000)*	*0.0018 [0.0003] (0.000)*	*0.0053 [0.0003] (0.000)*
INITIAL COST	*0.0003 [0.0027] (0.923)*	*0.0048 [0.0030] (0.112)*		
IMPLPHASE	*-1.3812 [0.0389] (0.000)*	*0.0511 [0.0418] (0.221)*	*-1.3557 [0.0381] (0.000)*	*0.0482 [0.0333] (0.149)*
DRRU	*42.0435 [4.3663] (0.000)*	*38.8320 [4.6669] (0.000)*	*41.5236 [4.2510] (0.000)*	*37.9306 [3.6947] (0.000)*
DCSPP	*21.3950 [4.8514] (0.000)*	*12.6426 [5.1749] (0.015)*	*21.2230 [4.7542] (0.000)*	*13.3243 [4.1366] (0.001)*
DTA	*126.6826 [6.6537] (0.000)*	*-20.3786 [7.6012] (0.007)*	*127.7805 [6.5350] (0.000)*	*-21.3393 [6.0619] (0.000)*
DSTATES	*-6.7338 [5.9723] (0.260)*	*9.6266 [6.3671] (0.131)*	*-7.0454 [5.8716] (0.230)*	*11.3310 [5.0564] (0.025)*
DMRICH	*-3.5019 [4.3897] (0.425)*	*1.7691 [4.6549] (0.704)*	*-4.4538 [4.3183] (0.303)*	*1.6225 [3.7226] (0.663)*
DRICH	*-7.1624 [4.5287] (0.114)*	*-1.3236 [4.7965] (0.783)*	*-7.1073 [4.4368] (0.110)*	*-0.8019 [3.8363] (0.834)*
DNE	*2.1731 [7.2072] (0.763)*	*14.7062 [7.6990] (0.056)*	*2.3591 [7.0501] (0.738)*	*16.0438 [6.1209] (0.009)*
CONSTANT	*404.0795 [22.5214] (0.000)*	*490.4227 [23.8009] (0.000)*	*396.2734 [22.1004] (0.000)*	*490.8609 [19.0289] (0.000)*
Observations	*928*	*928*	*928*	*928*
Pseudo R2	*0.1851*	*0.2143*	*0.1851*	*0.2128*

** Robust standard error in square parentheses. P-value in round parentheses.*

Road versus Railways Projects:

	OLS Regression		Quantile Regression	
	COSTOVERRUN (%age)	TIMEOVERRUN (%age)	COSTOVERRUN (%age)	TIMEOVERRUN (%age)
PCGETIMEOVERRUN	0.2111 [0.0682] (0.0022)		0.0748 [0.0397] (0.0608)	
TIMELAPSE	-2.661 [0.8934] (0.0032)	-1.5371 [0.4691] (0.0012)	-4.313 [0.3300] (0.0000)	-1.3811 [0.2834] (0.0000)
TIMELAPSESq	0.0042 [0.0015] (0.0070)	0.0016 [0.0008] (0.0440)	0.007 [0.0006] (0.0000)	0.0014 [0.0006] (0.0145)
INITIALCOST	-0.0284 [0.0245] (0.2472)	0.0511 [0.0263] (0.0534)	-0.0132 [0.0263] (0.6146)	0.0456 [0.0227] (0.0453)
IMPPHASE	0.6902 [0.2075] (0.0010)	-2.13 [0.2304] (0.0000)	0.5561 [0.1995] (0.0057)	-1.704 [0.1443] (0.0000)
DMRICH	4.4121 [6.8747] (0.5217)	-22.7683 [7.0347] (0.0014)	1.0267 [8.9337] (0.9086)	-10.0423 [7.8597] (0.2024)
DRICH	3.3794 [5.4331] (0.5346)	-8.6586 [6.2490] (0.1673)	1.8697 [8.1228] (0.8181)	-5.2196 [7.0547] (0.4600)
DNE			18.7574 [20.2666] (0.3555)	2.9748 [16.5987] (0.8579)
DRAILWAYS	-7.974 [7.9671] (0.3180)	42.2872 [7.9934] (0.0000)	-0.8883 [9.3754] (0.9246)	34.0317 [7.8898] (0.0000)
Constant	386.6825 [130.8940] (0.0035)	455.8811 [72.7784] (0.0000)	635.6869 [52.9052] (0.0000)	409.2738 [41.6311] (0.0000)
Observations	229	229	292	292
R-squared	0.41	0.56	0.2363	0.2896