# Comment: 'Procurement Choices and Infrastructure Costs'

*S. Vasserman*

It is quite clear from the accompanying article that not only are procurement practices ubiquitous and highly impactful for society, but they are also highly variable and comparatively understudied. The latter two conditions go hand in hand: there is such variety in conditions, restrictions and requirements across procurement projects that the vast theoretical literature on contracts and auctions (and, in fact, on models of procurement specifically) fails to generate many easily applicable rules of thumb.

A notable exception is "Auctions and Negotiations" by Jeremy Bulow and Paul Klemperer (1996), which has become known as a "critique" of efforts to optimize auctions with clever incentive schemes rather than simply solicit the entry of additional competitors. Even this observation, however, relies on key modeling assumptions that rarely hold up in practice—complete contracts key among them—without which, the validity of its conclusions is difficult to determine. What this means is that purely theoretical work is not enough: we need applied empirical work that can simultaneously speak to the validity of modeling assumptions in each setting, and assess concrete policies that have been undertaken or proposed to the maximal extent possible.

Why then, are there so few empirical studies? The article suggests a clear answer:

> *To make progress on this front we need to introduce **systematic** infrastructure cost and performance **benchmarking**, which will also **include procurement choices** as an explanatory variable. This point however is not yet recognized by policy makers.*

Why benchmarking, why include procurement choices, and what does it mean for all of this to be systematic? The nature of policy analysis is comparison. We wish to know how an existing procurement practice compares to a proposed alternative. As we can only ever try one variant at a time, we must rely on inferences of how an alternative *would* turn out on the basis of comparisons across existing examples. But comparisons require matching apples to apples. No two bridges are exactly alike, but a bridge in Massachusetts is similar to a bridge in Connecticut. Knowing precisely how similar they are is critical for determining how similar their construction costs and durations would be if conditions and procedures were the same – and if they were different.

There is no doubt that precisely *what* benchmarking standards should account for and include should be left for the experts to determine. But economists have a few requests that are worth listening to. It is generally insufficient to just record what work was done and how much was paid for it. In order to infer whether a project was done efficiently, there needs to be a record detailing how expectations for speed, cost and quality were formed prior to its completion. This record must be as comprehensive and objective as possible. Crucially, it must be designed to be interpretable by any knowledgeable researcher.

A procurement manager in rural England may not know why a given railroad track near Berlin required the numbers of hours, screws, steel slabs, etc. that it did or why certain numbers wound up exceeding initial projections. But with enough examples to compare against, he will be able to trace out patterns for which tracks are easier or cheaper to build and which features correspond to more efficiency or fewer surprises. Such insights are precisely what is needed to help determine the best procurement policy for each circumstance. For instance, variability between initial projections and ultimate usage suggests that there is contractual incompleteness. Consistency across many contractors is a sign of a competitive environment. The former observation suggests that careful negotiations may be preferable, while the latter suggests that barriers to eligibility for competing contractors may be unnecessarily prohibitive.

Moreover, these insights not only plug into established theoretical frameworks to suggest which rules of thumb are likely apply, but they offer a path to more extensive empirical research. Bajari, Houghton, and Tadelis (2014) measured the cost of contractual incompleteness in California road pavement projects and found that while clearly relevant, it did not constitute a major loss. De Silva, Kosmopoulou, and Lamarche (2009)

measured the impact of releasing internal cost estimates for highway construction projects in Oklahoma, and found substantial improvements in the competition between incumbent and entrant firms. Detailed empirical studies of this sort can offer not just rules of thumb, but also rigorous, tailored policy recommendations.

Impactful benchmarking is not impossible, but it requires explicit reporting requirements. It is no coincidence that the majority of existing empirical work on infrastructure procurement centers on projects with long-standing, detailed reporting standards like road paving and bridge construction. These types of projects are notably modular and formulaic, and so they are particularly suitable for comprehensive record keeping. But this is not the only reason for their unusual availability of detailed data. Absent explicit benchmarking standards, procurement agencies often keep only minimal records that are required for accounting purposes. When detailed records aren't explicitly required, even the accounting that *is* done is often difficult to locate afterward. As a consequence, only projects whose procurement procedures require detailed cost delineation for public accountability have systematic records available.

The selected availability of detailed procurement data greatly limits researchers' ability to assess competing procurement policies, but it is not inevitable. For example, Luo and Takahashi (2019) compare procurement projects by the Florida Department of Transportation (FDOT) that use either a "unit-price" (UP) format or a "fixed-price" (FP) format. Unit-price procurement requires a comprehensive bill of quantities that details the amount of each input that FDOT engineers estimate will be needed for construction. Contractors submit unit bids for each input item and are ultimately compensated based on the *actual* quantity used. As such, detailed information regarding both estimated and realized costs is kept for these projects. By contrast, fixed-price contracts require only a single price bid for the entire project and so no detailed delineation of costs and expectations is kept.

However, the distinction between projects that are procured with unit-price or fixed-price formats is hazy at best. As documented by Luo and Takashi, FDOT managers choose which format to use for each project on the basis of heuristic guidelines. Thus, there is no inherent reason why projects procured with fixed-price contracts could not have records of a similar specificity to those procured with unit-price contracts—had a unit-price format been chosen instead, the records would have been made by default. Understandably, organizations like FDOT simply choose not to expend the additional effort (and cost) to create detailed records if they are not mandated to do so. As a consequence, even comparisons between similar projects procured with fixed-price and unit-price procurement formats are very difficult. Although Luo and Takashi employ a number of clever econometric techniques to demonstrate the relationship between these formats in their paper, they are fundamentally limited in their ability to compare: there is no way to infer how a fixed-price contract would look in unit-price form.

It is precisely because different procurement projects are amenable to different forms of record keeping that standardized benchmarks are so important. Infrastructure projects come in all shapes and sizes; designing, planning, paying and building comes in many distinct forms. As such, it may not generally make sense to require item-level expectations and realizations for all inputs, as in the unit-price contract case. However, there is a tractable middle ground. In order to be maximally effective, benchmarks must be chosen to be *systematic* in the sense that they can be collected consistently across comparable infrastructure projects. For each category of project (perhaps grouped by scale and infrastructure type), there should be a set of meaningful metrics for uncertainty and performance of cost, speed, quality and competition. Examples of metrics may include the composition of generic (e.g. fuel, concrete) and specialized (e.g. specially molded parts) inputs, the relationship between inputs and time to completion, and the extent to which the scope and cost of the project is likely to depend on the contractor chosen and unforeseeable developments during construction. In addition to these features, it is important for benchmarks to include metrics of competition: the number of contractors who contend for a project, the extent to which the procurer can impose quality control and (crucially) the way in which the contractor is selected would go a long way.

Benchmarks don't need to be perfect. The variety of empirical papers discussed in the main article (as well as a number of others) demonstrate how researchers can creatively extrapolate from partial data and natural quasi-experiments to build an understanding of which procurement practices work well, when and why. But the more that better, more comprehensive data becomes available, the faster and more convincing that research will come out. This would be a welcome change.