SYM-AM-25-357

Excerpt from the Proceedings

of the

Twenty-Second Annual Acquisition Research Symposium and Innovation Summit

Volume III

Timing is Everything: Schedules, Models, and Analysis

Published: May 5, 2025

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School

Timing is Everything: Schedules, Models, and Analysis

Charles Pickar—retired from the Naval Postgraduate School (NPS) faculty, where he taught project management, defense acquisition, and systems engineering. Previously, he led the Applied Systems Engineering Program Area at Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry, including Director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. He has also served as Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focus on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@gmail.com]

Raymond Franck—retired from the NPS faculty in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active-duty career included operational tours, staff positions, leadership assignments, and, finally, head of the Department of Economics and Geography, United States Air Force Academy. His published work includes journal articles, book chapters, and research reports on military innovation and defense acquisition management. [cfranck215@aol.com]

Abstract

The complexity and scale of defense projects necessitate innovative project management and scheduling approaches. Digital twins, a digital representation of physical entities, transform how projects are planned, executed, and monitored. This paper explores the definition, applications, advantages, and challenges of implementing digital twins in project management. Additionally, integrating artificial intelligence (AI) predictive delay analysis processes provides an advanced framework for optimizing execution and risk mitigation. This paper examines (a) how real-time digital replicas and AI-driven predictive analytics using defense acquisition data can enhance decision-making, efficiency, and project outcomes in defense project management and (b) how prediction markets might enhance the timeliness and quality of information for program management—leading to better program outcomes. While the technical advances are impressive, they rely on information. Also, program management involves human skills and knowledge. Prediction markets have demonstrated promising capabilities to provide timely and accurate information for program management—with or without state-of-the-art technical means.

Digital Twins and Supporting Tools

Introduction

Project management requires tools and methods to plan, execute, and monitor progress. The advent of digital twins (DTs), technical advancements like artificial intelligence (AI), and techniques like predictive analytics offer modern alternatives to traditional tools.

Schedules outline the planned sequence and duration of weapon system development events. When accurate, they offer warfighters a dependable delivery date. However, programs often miss these deadlines, breaking warfighter trust and leading to cost and performance issues. Defense project management involves intricate planning, execution, and monitoring to meet stringent requirements. Traditional methodologies struggle to address unforeseen challenges, resource constraints, and delays effectively. A DT provides a virtual (and simulated) replica of a physical project (weapon system development), allowing project managers to simulate, analyze, predict, and execute outcomes in real time. The convergence of DTs, AI, and predictive delay analysis (PDA) offers an innovative paradigm shift, leveraging real-time acquisition data and predictive insights to enhance project execution.

This paper explores the feasibility of utilizing AI and DT approaches to offer a model for creating and executing project schedules. By leveraging these advanced technologies, the study aims to enhance accuracy, efficiency, and adaptability in project scheduling processes.

The rise of DTs and AI marks a new era for project management, particularly in schedule management and the potential for improving the schedule estimating and executing process. Instead of using virtual replicas of the system being developed, we suggest using a DT as a model or simulation of the developed weapon system. With AI's analytical power, this idea can improve project timelines, predict delays, and boost efficiency. DTs serve as the activity monitoring tool. PDA using the schedule delay factor data supported by acquisition data is incorporated into the AI. AI then monitors the DT's inputs and alerts the decision-maker to the need for action. Predicting schedule problems is the end state.

This effort is part of a continuing research agenda started by Franck et al. (2016). The latest paper in this line of research (Franck & Pickar*, 2024*) discussed complexities in the estimation and execution dynamic processes that determine success or failure in schedule execution and the "wisdom of crowds" and prediction markets concepts and relationship to schedule and program management decisions (Franck & Pickar, 2024).

Literature Review

This study explores using DTs as an environment for practicing project management decision-making. The tools we propose to assist in this effort are DTs as a simulation environment and PDA and machine learning (ML) as a subset of AI. The literature sources include *inter alia* defense-focused papers and systems research on managing defense projects. An unexpected resource for studying how DTs and delay analytics can be used in defense project management has been the construction industry, where practitioners have explored the use of DTs, PDA, and AI for some time.

Grieves initially proposed the concept of DTs in the context of Product Life Cycle Management (PLM) in 2002 (Grieves, 2002). DTs were originally digital replicas of physical assets. They have since evolved into models of non-physical systems and processes. Grieves and Vickers believe DTs offer a dynamic approach to managing the development of complex systems (Grieves & Vickers, 2017). In project management, DT models can replicate project components, facilitating improved planning, execution, and monitoring. To manage weapon system development execution, DTs can provide a real-time linkage between the work accomplished and the project management schedule (and plan), providing the PM insight into favorable and unfavorable execution developments. Figure 1 shows the basic model (modified to add the project management function; Grieves, 2002).

Figure 1. Project Management DT Environment
(Grieves, 2002)

DTs were initially used in the aerospace sector, where NASA used mirrored simulation models for system maintenance and failure prediction (Uhlenkamp et al., 2019). Over time, DTs have evolved to integrate the Internet of Things (IoT) sensors, AI, ML, and data analytics, providing real-time insights into the development of the system and enabling life cycle management (Attaran & Celik, 2023; Pan & Limao, 2021).

Project schedules are critical for completing projects, as they outline the timeline of tasks and activities. DT models can enhance project scheduling by providing real-time data and predictive analytics. Project managers can identify potential delays by simulating different scenarios and adjusting schedules accordingly (Tao et al., 2019). DTs can provide a data-driven decision-making tool by continuously analyzing project performance. This real-time monitoring provides better forecasting of project timelines and improved resource allocation by understanding workload distribution, enabling proactive real-time adjustments. A DT also simulates real-world scenarios for the project process environment, providing a safe environment for exploring solutions to project problems.

Predictive Delay Analysis and Machine Learning

ML and PDA are distinct yet complementary concepts within data-driven decision-making. ML is a subfield of AI that involves the development of algorithms capable of identifying patterns in data and making predictions or decisions without explicit programming (Mitchell, 1997). These algorithms, including regression models, decision trees, support vector machines, and neural networks, have been widely used in healthcare, finance, and manufacturing due to their flexibility and predictive power.

In contrast, PDA is a specialized technique within project management, predominantly in the construction and engineering sectors. Its primary objective is to forecast project delays by analyzing scheduling data—typically derived from critical path method (CPM) schedules—alongside historical and real-time performance metrics (Arditi & Pattanakitchamroon, 2006). PDA relies on various methods such as trend analysis, rule-based systems, Monte Carlo simulations, and, increasingly, ML techniques.

Although PDA is domain-specific, ML serves as the general-purpose analytical thread. The intersection of the two arises when ML is used to improve the accuracy and automation of delay predictions. For example, supervised ML models can be trained on historical project data to identify patterns in activity durations, programmatic issues, procurement timelines, or subcontractor performance indicative of future delays.

The difference between PDA and ML lies in their scope and application. While ML offers a broad methodological framework applicable to numerous domains, PDA is a targeted application for temporal risk and delay forecasting in projects. Integrating ML into predictive delay analysis represents a convergence of generalizable algorithmic intelligence with domain-specific scheduling insights, potentially yielding more data-informed project risk assessments (Marzouk & El-Rasas, 2014).

Classical scheduling methods require manual estimation and static modeling of task durations and dependencies. These approaches lack the flexibility to respond to real-time changes and uncertainties, leading to inaccurate timelines and inefficient resource use (Kerzner, 2022). Leveraging ML and AI, recent studies demonstrate a growing shift from traditional reactive approaches to predictive, data-driven methodologies. PDA traces its origins to traditional project scheduling and risk management methodologies, including Critical Path Method (CPM) and Earned Value Management (EVM). A PDA model typically uses ML, statistical modeling, or simulation to identify the probability and impact of delays before they occur.

Table 1. Core Components of PDA

Component	Description
Input Data	Project schedules, progress reports, weather data, resource logs, risk registers, subcontractor performance, etc.
Features	Task duration, lag times, critical path activities, manpower fluctuations, change orders, etc.
Modeling Technique	Regression, Decision Trees, Random Forests, Bayesian models, or Deep Learning for complex dependencies
Output	Probability of delay per task, delay forecasts, and risk classification (low/medium/high)
Visualization	Gantt overlays, risk heatmaps, delay likelihood timelines

Table 1 lists the components of PDA. Of note are the input and output components, which will be included in the model suggested later in this paper.

A PDA model comprises the systematic collection and integration of data, which consists of historical schedule performance metrics, explicitly contrasting Planned versus Actual outcomes, and documentation of change orders and incidents of rework. The appropriate analytical technique is determined upon the compilation and preparation of this data. Potential methodologies include logistic regression for estimating the likelihood of delays or applying random forest algorithms to enhance interpretability and address nonlinear relationships (Ghimire & Mishra, 2019). The deployment of the delay analysis model is intended for integration within contractor-available project management software platforms, such as Primavera, MS Project, or JIRA, with the capacity for real-time updates through the incorporation of progress feeds throughout the execution phase. Different established visualization tools, including Power BI and Tableau, could be utilized to present real-time forecasts and associated risks to the critical path. Potential use cases include detecting likely delays in hardware delivery due to supplier risk and forecasting design freeze violations on critical components. Using feedback loops from the DT to auto-update the schedule would keep the PM in the loop (Liu et al., 2021).

PDA supports proactive project management by providing an opportunity to anticipate potential delays (Lee, 2017). It can also improve the accuracy of project timelines, making it easier to set realistic deadlines. The awareness of when and where delays might occur allows managers to allocate resources more effectively to avoid chokepoints.

Gondia et al. examined the use of supervised ML algorithms—including Support Vector Machines, Decision Trees, and Random Forests—to predict construction project delays by modeling the intricate interdependencies of delay risk sources (Gondia et al., 2020). The paper finds that multifactorial and interrelated risks often cause construction delays that traditional methods struggle to predict accurately. The study further reveals that ML models can effectively capture nonlinear relationships between variables and deliver more reliable and proactive delay risk assessments. Notably, the Random Forest algorithm demonstrated superior predictive accuracy and interpretability performance.

Machine Learning Algorithms in Classification and Regression

ML models have proven indispensable tools in modern PDA. These models, particularly those employing supervised learning algorithms, can analyze vast amounts of data to identify patterns and accurately predict outcomes. Support Vector Machines (SVMs), Decision Trees, and Random Forests are the most notable algorithms used for this purpose. Each algorithm presents unique characteristics regarding modeling capacity, interpretability, and performance.

Support Vector Machines

SVMs are supervised learning models primarily employed for classification tasks, although they can also be adapted for regression problems. The core principle of SVMs is to determine an optimal hyperplane that separates data points of different classes with the maximum possible margin. The data points closest to the decision boundary, termed *support vectors*, are critical in defining this margin (Cortes & Vapnik, 1995).

SVMs are particularly effective in high-dimensional spaces and are well-suited to problems where the number of features exceeds the number of observations. In cases where the data is not linearly separable, the SVM utilizes a *kernel trick* to project the data into a higher-dimensional space where a linear separation becomes feasible (Schölkopf et al., 1998). Common kernels include linear, polynomial, radial basis function (RBF), and sigmoid.

Decision Trees

Decision Trees are versatile, non-parametric models capable of performing both classification and regression tasks. They operate by recursively partitioning the input space into subsets based on feature values, forming a tree-like structure of decision rules (Quinlan, 1986). At each internal node, the algorithm selects a feature and a corresponding threshold that best splits the data to increase *purity*—a measure of how homogenous the resulting subsets are to the target variable.

The interpretability of decision trees is a significant advantage, as the resulting model can be easily visualized and understood. However, decision trees are prone to overfitting, especially when they grow deep and complex. Techniques such as pruning, setting maximum depth, or limiting the number of samples per leaf are commonly used to mitigate overfitting (Loh, 2011).

Random Forests

Random Forests is a learning method that enhances the performance of decision trees by constructing a multitude of trees and aggregating their outputs (forest). This approach, known as *bagging* (bootstrap aggregating), involves training each tree on a random subset of the data and selecting random subsets of features at each split, thereby introducing diversity among the trees (Breiman, 2001).

For classification tasks, the final output is determined by a majority vote among the individual trees, whereas for regression, the output is the average of the predictions. Random Forests offer improved generalization performance over individual decision trees, primarily due to their ability to reduce variance and avoid overfitting. Although Random Forests are less interpretable than single decision trees, they are generally more accurate and robust across various tasks. The trade-off lies in increased computational requirements and reduced transparency. These algorithms form a foundational component of many ML pipelines and continue to be extensively applied across various fields, including bioinformatics, finance, and engineering applications.

Table 2. ML Algorithm Summary Comparison

Algorithm	Application	Strengths	Limitations
SVM	Classification / Regression	Effective in high-dimensional spaces; robust with clear margin separation	Computationally intensive; sensitive to parameter tuning
Decision Trees	Classification / Regression	Simple to understand and interpret; fast training	Prone to overfitting; high variance
Random Forests	Classification / Regression	High accuracy; reduces overfitting; robust to noise	Less interpretable; increased training time

Key challenges for all ML algorithms include poor data quality, lack of project standardization, and resistance to adopting opaque "black-box" models. Many ML models require large, well-structured datasets for training—something not all organizations possess. Furthermore, the lack of transparency in complex models such as neural networks hinders decision-maker adoption (Barbierato & Gatti, 2024).

Application to Weapon System Development

Weapon system development projects are characterized by long durations, high uncertainty, and tight integration of subsystems—making them particularly susceptible to cascading delays. PDA offers an opportunity to shift from reactive project recovery to proactive risk management.

ML models can detect early signals of schedule slippage by integrating real-time data from testing cycles, supplier schedules, and subsystem integration reports. This approach mirrors that used by Awada et al. (2021), where field data integration enhanced mid-project forecasting (Awada et al., 2021).

Furthermore, hybrid approaches can enable defense agencies to tailor models using historical program data and expert judgment (Fitzsimmons et al., 2022). This would allow nuanced modeling of risk factors such as integration complexity, geopolitical disruptions, and budgetary constraints. PDA and ML can transform schedule management in weapon system development—enabling agile responses to risk, optimizing resource allocation, and increasing the likelihood of on-time, on-budget delivery of critical defense capabilities.

Acquisition Data

DTs, PDA, and ML require foundational historical data. The Department of Defense (DoD) collects data throughout the development process, including databases maintained by OSD and CADE. The intense DoD focus on cost dictates the data collected and its format. At the risk of adding more work, we suggest a review to collect and format data that can support schedule-focused PDA. ML could better predict schedule problems and address cost issues resulting from schedule delays.

As noted above, delay analysis requires that delay data be integrated into the model or the DT. Scheduling necessitates an analysis of the factors that have historically contributed to prolonged development times. Over the past 20 years, extensive research has identified several contributing elements to schedule delays, (Drezner & Smith, 1990; Van Atta et al., 2015). These factors include budget constraints, funding issues, complexity, technical challenges, and requirements. Building on these studies, research performed in 2018 used Selected Acquisition

Reports (SAR) from the OSD acquisition databases to identify delay factors cited by project managers during their annual SAR submissions (Pickar, 2018). Table 3 is a list of schedule delay factors developed in that study.

Table 3. Schedule Delay Factors
(Pickar, 2018)

Schedule Delay Factor
Administrative changes to schedule include updates to the APB and ADM as well as changes resulting from Nunn-McCurdy processes and program restructuring
Technical
Testing delays
Delay in the availability of key capabilities/ facilities (launch vehicle/ testing facilities/ IOT&E units)
Budget/ Funding Delays
Delays attributed to the Contractor
Delays because of Rework
External events such as inflation, earthquakes, labor strikes, etc. *(Force Majeure)*
Delays due to Contracting/ Contract Negotiation/ Award delays

Figure 2 is the proposed DT planning and execution model, with the planning aspect highlighted in red. The DT model is initially used for planning as part of the planning process. The work breakdown structure and task duration estimates are included in the input data (shown as data sources). The initial schedule data is the raw CPM-derived schedule, which is the basis for inputting data into the model. A key part of identifying and preparing the acquisition data is filtering by type of system (aircraft, missile, etc.). Schedule Delay Factors with frequencies are compared to the WBS and other planning data to determine similar occurrences. Once the project plan is built, the AI enhancements provide the necessary data-driven adjustments and a risk estimate. Outputs are the completed schedule and initial identification of schedule risk. Other risk areas are also identified.

Figure 2. Planning Aspects of the DT Approach

Figure 3 is the complete model incorporating the planning aspects. The diagram illustrates a comprehensive framework for integrating the DT model into the weapon system development project management processes, focusing on enhancing schedule optimization and decision-making in complex projects. It outlines the dynamic interaction between data acquisition, core processes, AI enhancements, and decision support mechanisms.

Figure 3. DT Planning and Execution Model

The process starts with acquisition data, which feeds into the input data stage. As noted above, the acquisition data currently used for this study is from the DoD SAR. As data availability on active programs increases throughout the DoD, the available data will increase exponentially, improving the process. This stage consolidates various types of project-related data, including project schedule data, schedule delay factors, and other relevant data sources. The DT then utilizes these inputs, a virtual representation of the physical project environment that enables simulation and analysis in real-time.

The core processes, comprising planning and execution activities, directly interact with the DT. This interaction ensures that the simulation model remains synchronized with actual project developments and strategic planning efforts. The DT generates key outputs such as an optimized schedule, risk mitigation strategies, and resource forecasting. These outputs are critical for effective project management and feed into the decision support system. This system offers dashboard insights, automated alerts, and proactive recommendations to support informed and timely decision-making by project stakeholders. An essential part of the model are the feedback loops throughout. As information is received, the model can be continuously updated to improve the execution.

Enhancing this entire framework are the AI enhancements, which integrate advanced analytical capabilities, including PDA, AI-driven schedule adjustments, and resource allocation optimization. These AI modules provide feedback to the DT, allowing for iterative improvements and adaptive project planning.

The diagram encapsulates a closed-loop, intelligent project management ecosystem where a DT is the central analytical engine. It continuously ingests data, interacts with planning processes, and outputs actionable insights, all enhanced by AI-driven modules to optimize performance and reduce project risks.

Concluding Thoughts on Digital Twins

Integrating DT technology into project management systems allows organizations to monitor, simulate, and optimize projects in real-time. It enhances risk management, decision-making, collaboration, and efficiency, making project execution more predictable and effective.

Adopting DTs in project management changes how projects are planned, executed, and monitored. By offering real-time insights, predictive capabilities, and enhanced collaboration, DTs can address traditional project management challenges.

Future research should focus on standardizing DT implementations and exploring their integration with emerging technologies like blockchain and quantum computing.

Prediction Markets: Synthesizing Scattered Information

Introduction

“Executives know . . . valuable information is scattered across the organization. They just don’t know how to retrieve it” (Thompson, 2012, p. 1).

“Those who made that (Challenger launch) decision were unaware of the recent history of problems concerning the O-rings” (Rogers Report, 1986, p. 88).

Our joint line of inquiry is focused on using relevant information to manage acquisition programs more effectively. However, data-driven actions are unlikely to be more efficacious than quality of the data, as noted above. Effective program management is more likely to be achieved through information derived from various sources and methods. In this part of our paper, we consider information derived from the program team members—suppliers and DoD program managers. As experience shows (e.g., the Challenger mishap), failure to make informed decisions (in acquisition and operations) is much easier said than done.

Along these lines, much information resides inside an organization (Thompson, 2012), but obtaining high-quality information through regular channels has often proved difficult. But well-organized markets have shown potential for eliciting information (Hayek, 1945, pp. 17, 19–23). In this context, prediction markets are a relatively recent method to elicit useful information—potentially valuable for defense acquisition managers.

The following sections focus on actual cases of markets’ (including prediction markets’) potential for aggregating information useful for acquisition program decisions. We also identify some potential problems for prediction markets in a defense acquisition context.

Identifying the Challenger Accident's Causes

The stock market's response to the Challenger Space Shuttle loss on January 28, 1986 demonstrated a well-organized market's ability to gather and sift information. Challenger launched its 10th mission. Shortly after getting airborne, the Shuttle experienced a catastrophic failure of its booster rockets. The Challenger was destroyed, with the loss of all crew members.

President Reagan directed a major investigation conducted by a select commission—chaired by former Attorney General William Rogers. The Commission was formed on February 6, issued its report on June 6 (Rogers Commission, 1986, pp. i, iii), and found that failure of O-rings intended to keep rocket engine thrust properly contained was the sole (proximate) cause of the accident. The report (Rogers Commission, 1986, p. 45) stated this conclusion:

> the loss of the Space Shuttle Challenger was caused by a failure in the joint between the two lower segments of the right Solid Rocket Motor. The specific failure was the destruction of the seals that are intended to prevent hot gases from leaking through the joint during the propellant burn of the rocket motor.

By way of root causes, the Rogers Commission (1986, p. 88) noted the following. The launch decision "was flawed. . . . Those who made that decision were unaware of the recent history of problems concerning the O-rings . . . *If the decision-makers had known all of the facts*, it is highly unlikely that they would have decided to launch" (Rogers Commission, p. 88, emphasis added).

While the Rogers Commission conducted its investigations and deliberations, stock market investors incorporated information on the Challenger's loss—concerned, *inter alia*, with the adverse effects on the company's stock price, whose product was the proximate cause of that loss. As Maloney and Mulherin (2003, p. 453) reported, "in the period immediately following the crash, securities trading in the four main shuttle contractors singled out the proximate cause of the accident (indirectly) by identifying the firm that manufactured the faulty component."

We note two significant features of this story. First, while the infamous O-rings received more publicity, the failure to make a properly informed launch decision was more consequential. Second, the Rogers Commision (1986, p. 88) makes clear that risks associated with colder-weather launches was available but somehow did not find its way to those who actually chose to launch on January 28. This is impressive support for Thompson's (2012) assertion cited above.

However, there is fairly strong evidence that interested participants in the stock market quickly focused on those O-rings and behaved accordingly. In fact, there was a suspension in trading of Thiokol[1] shares for a period of time (Maloney & Mulherin, 2003, p. 453). In short, the marketplace, on this occasion, performed according to Hayek's contention that markets are highly effective processors and aggregators of decentralized information.

Could a well-designed prediction market have gotten the relevant facts to those decision-makers? Observable events strongly suggest that a stock market could sort out the causes of the Challenger disaster—albeit *ex post*.

The 2024 Presidential Election: Prediction Markets vs. the Pollsters

During that election, the standard polls and "experts" predicted a very close race (270 to Win, 2024). One forecasting exercise gave Harris a 50.12% chance of winning (based on an extensive simulation exercise; Osipovich, 2024a). As it turned out, Trump won handily: electoral college, popular vote, and the "blue wall" states. He won all his "safe" states and also won many "contested" states (270 to Win, 2024; NBC Chicago, 2024).

[1] The manufacturer of those O-rings.

In this case, the prediction markets did better than the pollsters (Ferguson & Rincon-Cruz, 2024). One major participant (the "Trump Whale") turned out to be prescient: picking a Trump win in the electoral college, the popular vote nationwide, and the primary battleground states (Osipovich, 2024c).

Inquiries as to why this happened led to multiple hypotheses. One was a right-wing conspiracy—intended to create a Trump bandwagon effect and discourage potential Democratic voters (Osipovich, 2024c). That the Trump Whale turned out to be a non-U.S. citizen who claimed to be interested in making money (Osipovich, 2024c) cast doubt on that particular hypothesis.

A second hypothesis was that prediction markets were better suited for forecasting this election (and probably others). In our opinion, the rationale centers on incentives:

- Those intending to vote for Trump were said to be shy about admitting it (Osipovich, 2024c)—possibly due to his widespread vilification (e.g., Fitton, 2025) before and throughout the campaign.
- Getting a representative sample through traditional polling methods has become more difficult for several reasons. Those inclined to participate in a survey likely reflect a sort of selection bias.[2] That is, willingness to participate in a polling survey was correlated with their propensity to vote for a particular candidate. In this context, the shy-Trump-voter hypothesis could well have significant explanatory power (Osipovich, 2024c).
- The most serious allegation against the polling establishment was tailoring its products to the interests of major news sources. The rationale is that too-close-to-call polling reports are more newsworthy, and therefore, the business interests of the polling agencies were more closely aligned with the too-close-to-call reporting customer interests than identifying the Trump advantage (Osipovich, 2024c).

There are good reasons to believe this particular case reveals a great deal.

First, the prediction markets were indeed better at estimating election results than the regular polls. This is consistent with the hypothesis that prediction markets are often better than "experts" in generating information (Ferguson & Rincon-Cruz, 2024).

Second, incentives matter, and prediction markets provide (at least) pretty good incentives to be right. And there is a good reason that prediction markets work well even with relatively small stakes (Servan-Schreiber, 2004; Yeh, 2006).

Carmageddon

From July 15 to 18, 2011, a stretch of Interstate 405 in the Los Angeles area was closed for improvements. The associated traffic problem was confidently expected to be severe. As one observer put it, "I feel our collective psyches might not be able to withstand a traffic jam of this magnitude" (Gostar, 2011). In the event, however, traffic in the area in question was significantly lighter than normal, and the massive traffic jam simply didn't happen.

Why was this so? A good general answer is self-negating prediction. More specifically, the closure was announced well in advance and motorists were well informed about construction schedules and alternative routes. It appeared that the reason for the unexpectedly light traffic was motorists' behavior—likely based on trust in the consensus prediction.

- They stayed home as advised (Gish, 2011); more likely, they rescheduled trips so as to be at home during the expected congestion.

[2] In part due to widespread ability to screen incoming telephone calls.

- Information about alternate routes kept traffic away from the construction area.
- The availability of public transportation, such as the Metrolink commuter rail system, kept some potential motorists off the highways.[3]

In short, motorists believed the traffic congestion forecasts and took steps to avoid the situation. As a result, traffic difficulties were much less than the consensus prediction.

If there had been a prediction market about the size of the traffic snarl, it's possible the market outcome would have been a high probability of significant congestion. It's also likely that market participants would have gotten wind of motorists' plans and reflected the new information in their bets.[4]

The "Circularity" Issue

Along these lines, the "Trump Whale" in the 2024 election prediction market activity was a most interesting development. "Theo" (a pseudonym) placed several large bets on Trump's success in multiple markets using multiple account identifiers (Osipovich, 2024a).

Initially, there was some concern from the establishments (both polling and political) that Theo intended to create a synthetic groundswell of support for Trump's candidacy and engender a band-wagon effect that would improve his chances of being elected. "Theo," however, stated that he was in the prediction markets simply to make money—as discussed above (Osipovich, 2024b). Among other things, he commissioned polls whose core question was how the subjects' neighbors would vote—hypothesizing that Trump voters were reluctant to reveal their intentions, given the size and ferocity of the anti-Trump movement. Theo reported that this approach yielded more accurate information than standard polling methods.[5]

What seems particularly important, however, is that high-stakes prediction markets can create perverse incentives. Suppose, for example, that some "whale" places bets in a high-stakes terrorism prediction market—and is interested in making lots of money. Given the stakes, it is conceivably possible for someone to commission an act of terrorism in keeping with his position in the market. If a "whale" could contract with anonymous pollsters for information on voters' intentions, then contracting with a terrorist organization for an event that made his bet correct could be feasible—with significant financial returns for both parties. It's also conceivable that an agent of a terrorist organization could participate in a prediction market—seeking to make a terrorist event also result in financial benefits.

There is reason to conclude that market results affecting real-world outcomes is a major concern. For example, prediction market results indicated a significant increase in the Bush Administration's election chances in 2004 if Osama Bin Laden were killed or captured prior to November of that year (Thompson, 2012, p. 53)[6]. There is some possibility of prediction market results influencing the actions of decision-makers using market results to influence events.

This gets us to the more general issue of "circularity."[7] As generally organized, prediction markets generally assume (even if tacitly) that the event under consideration is not affected by the market's operation. However, the state of the market itself (generally as the probability of Event A) is open to inspection—to attract potential betters, if nothing else.

[3] For example, the Metrolink (a commuter rail system) set records for ridership during the construction period (The Source, 2011).
[4] Sometimes called "Bayesian Updating."
[5] Since Theo has, at this time, not disclosed the details of the polling he commissioned, one cannot verify the polls' results.
[6] There is, however, we don't know of any indication that market outcome influenced counter-terrorism operations in 2004.
[7] "Self-fulfilling" and "self-negating" beliefs are similar concepts. These are part of standard economic discourse.

This raises some interesting questions. At a minimum, the market provides information to decision-makers who have some control over Event A, such as when or if the event occurs. If the market results are deemed useful, affected decision-makers will likely include the prediction market's equilibrium probability in their calculations. This has been called the "circularity" problem (Thompson, 2012, pp. 145–6). Actors in the situation (inside or outside of the market) could take actions to affect events to further their own interests.

Suppose further that the prediction market offers a (equilibrium) probability for Event A. Suppose also that probability is conditioned on some other event, B. Suppose finally that agents who could control the probability of Event B are aware of the market results. They now have an incentive to take steps to increase (or decrease) the likelihood of B. One prediction market reported a substantial increase for President Bush's 2004 reelection if Osama bin Laden were captured. As noted above, this could have provided a significant incentive to increase efforts to apprehend bin Laden (Thompson, 2012, p. 53).[8]

More broadly, it's well known that beliefs about an event can affect the likelihood of the event occurring. The closure of a stretch of Interstate 405 in the Los Angeles area was widely publicized in advance as an extreme case of traffic congestion. In the event, however, traffic was lighter than usual, with transit times also less than normal. One reason cited was the shift to public rail transportation (The Source, 2011). The Carmageddon forecast indicates timely information could cause management to take actions that would avoid untoward developments (or mitigate them). That is, self-negating predictions are highly desired results in prediction markets in a program management context. If so, how do the market authorities determine the winners? And if they can't, then how does the market incentivize the participants?

Perhaps cleverly posed questions could solve (or lessen) this problem. If so, what would be the source for those questions, and by what means would they be "admitted" to the prediction marketplace?[9]

However, another question is whether a prediction market could provide helpful, timely information regarding launch temperatures' effects on the O-rings. Even if the proper questions were proposed before the fatal Challenger launch, could a prediction market have responded to provide timely information?

Concluding Thoughts on Information Gathering

For standard prediction markets, the circularity issue is a potential problem. Circularity can be a primary goal for prediction markets devoted to defense acquisition programs. If the current policy is highly likely to cause a significant, untoward result, then program management can perhaps take action to avoid or mitigate that result. The problem then becomes sorting out who placed the right bet. If not addressed, such possibilities substantially reduce the propensity for serious participation in the prediction markets. This seems a thorny issue, with some potential for entering a "wilderness of mirrors."[10]

While it's possible that well-designed betting propositions can avoid (or lessen) this problem, doing that would likely take some creative thinking.

More fundamentally, there are remaining questions regarding the potential of prediction markets to better inform program management decisions. This suggests further inquiries, including those we identify here. Since well-organized markets are generally more effective:

---

[8] However, we know of no indication that the Administration was influenced by this result.

[9] For example, PredictIt, a political prediction market site, had 13 U.S. markets in session on March 19, 2025. No indications of how to open a market.

[10] A popular (and descriptive) term in the intelligence literature—in both its factual (e.g., Martin, 1980) and fictional manifestations.

getting a better focus on the rules and customs of prediction market operations—to include, perhaps, field interviews; and conducting "gaming" studies of acquisition-oriented prediction markets in an experimental setting.

References

270 to Win. (2024). *Compare the 2024 presidential consensus and polling maps*. https://www.270towin.com/

Arditi, D., & Pattanakitchamroon, T. (2006). Selecting a delay analysis method in resolving construction claims. *International Journal of project management*, *24*(2), 145–155. https://www.sciencedirect.com/science/article/pii/S0263786305000906

Attaran, M., & Celik, B. G. (2023). Digital twin: Benefits, use cases, challenges, and opportunities. *Decision Analytics Journal*, *6*, 100165. https://doi.org/10.1016/j.dajour.2023.100165

Awada, M., Srour, F. J., & Srour, I. M. (2021). Data-driven machine learning approach to integrate field submittals in project scheduling. *Journal of Management in Engineering*, *37*(1), 42. https://doi.org/10.1061/(asce)me.1943-5479.0000873

Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, *13*(2), 416. https://doi.org/10.3390/electronics13020416

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32. https://doi.org/10.1023/a:1010933404324

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273–297. https://doi.org/10.1007/Bf00994018

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules.* RAND. http://scholar.google.com/scholar?q=related:Su8l_4wFOQ8J:scholar.google.com/&hl=en&num=20&as_sdt=0,5

Ferguson, N. & Cruz, M. R. (2024, November 16). How the Trump whale and prediction markets beat the pollsters in 2024. *Wall Street Journal*. https://www.wsj.com/politics/elections/how-the-trump-whale-and-prediction-markets-beat-the-pollsters-in-2024-dd11ec4e

Fitton, T. (2025, January 3). President Trump needs to clean house. *The Verdict*.

Fitzsimmons, J. P., Lu, R., Hong, Y., & Brilakis, I. (2022). Construction schedule risk analysis—A hybrid machine learning approach. *Journal of Information Technology in Construction*, *27*, 70–93. https://doi.org/10.36680/j.itcon.2022.004

Franck, R., & Pickar, C. (2024). *Navigating the labyrinth: Unraveling schedule complexity*. https://dair.nps.edu/handle/123456789/5114

Ghimire, S., & Mishra, A. K. (2019). Comparative study of prospective delay analysis techniques (DATs). *Saudi Journal of Civil Engineering*, *03*(05), 84–98. https://doi.org/10.36348/sjce.2019.v03i05.001

Gish, J. (2011). *I-405 summer weekend closure planned*. Inside Seven (CALTRANS). https://web.archive.org/web/20150912102517/http:/www.dot.ca.gov/dist07/Publications/Inside7/story.php?id=651

Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, *146*(1), 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

Gostar, R. & Burreson, S. (2011). Carpocalypse: The weekend the 405 freeway will stand still. *Brentwood Patch*. https://web.archive.org/web/20120718034543/http:/brentwood.patch.com/articles/carpocalypse-the-weekend-the-405-freeway-will-stand-still

Grieves, M. (2002). *Completing the cycle: Using PLM information in the sales and service functions [slides]*. In *SME Management Forum*.

Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In *Transdisciplinary Perspectives on Complex Systems* (pp. 85–113). Springer International Publishing. https://doi.org/10.1007/978-3-319-38756-7_4

Hayek, F. A. (1945). American economic review, (vol 35), 17–31. (Reprinted from *Price theory: Penguin modern economics readings,* by H. Townsend, Ed., 1971.).

Kerzner, H. R. (2022). *Project management: A systems approach to planning, scheduling, and controlling* (13th ed.). John Wiley & Sons.

Lee, B. (2017). *Proactive project management*. University of Chicago Press.

Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*, *58*, 346–361. https://doi.org/10.1016/j.jmsy.2020.06.017

Loh, W. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

Martin, D. C. (1980). *Wilderness of mirrors*. Ballantine Books.

Marzouk, M. M., & El-Rasas, T. I. (2014). Analyzing delay causes in Egyptian construction projects. *J Adv Res*, *5*(1), 49–55. https://doi.org/10.1016/j.jare.2012.11.005

NBC Chicago. (2024, November 5) What are the “blue wall” states?. https://www.nbcchicago.com/news/local/what-are-the-blue-wall-states-what-to-know-as-results-come-in/3593443/

Osipovich, A. (2024a, October 18). A mystery $30 million wave of pro-Trump bets has moved a popular prediction market. *Wall Street Journal*. https://www.wsj.com/finance/betting-election-pro-trump-ad74aa71

Osipovich, A. (2024b, November 1). The Trump-betting whale speaks out: “I have absolutely no political agenda”: Self-described French trader says he bet more than $30 million US election after analyzing polls.

Osipovich, A. (2024c, November 6). How the Trump whale correctly called the election. *Wall Street Journal*. https://www.wsj.com/finance/how-the-trump-whale-correctly-called-the-election-cb7eef1d

Pan, Y., & Limao, Z. (2021). A BIM-data mining integrated digital twin framework for advanced project management. *Automation in Construction*, *124*, 103564. https://doi.org/10.1016/j.autcon.2021.103564

Pickar, C. K. (2018). *Informing DoD program planning through the examination of the causes of delays in acquisition using acquisition data*.

https://calhoun.nps.edu/bitstream/handle/10945/58762/SYM-AM-18-079-019_Pickar.pdf?sequence=1

PredictIt. (2025). https://www.predictit.org/markets

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*, 81–106. https://doi.org/10.1007/BF00116251

Rogers Commission. (1986). *Report to the president on the space shuttle Challenger accident*. https://sma.nasa.gov/SignificantIncidents/ass

Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*(5), 1299–1319. https://doi.org/10.1162/089976698300017467

Servan-Schreiber, E. et al. (2004, September). Prediction markets: Does money matter?. *Electronic Markets 14*(3), 9–10.

Silver, N. (2024, October 7). *What's behind Trump's surge in prediction markets?.* https://www.natesilver.net/p/whats-behind-trumps-surge-in-prediction

Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics*, *15*(4), 2405–2415. https://doi.org/10.1109/tii.2018.2873186\

The Source (2011), *"Metrolink sets weekend ridership recording during 405 closure".*

Thompson, D. N. (2012). *Oracles: How prediction markets turn employees into visionaries*. Harvard Business Review Press, p. 1.

Uhlenkamp, J.-F., Karl, H., Wellsandt, S., Thoben, K.-D. (2019). *Digital twin applications: A first systemization of their dimensions*. In (p. 48). IEEE. http://dx.doi.org/10.1109/ice.2019.8792579

Van Atta, R. H., Kneece Jr, R., Patterson, C. M., & Hermes, A. C. (2015). Assessing weapon system acquisition cycle times: Setting program schedules. *Institute for Defense Analysis IDA*. http://www.dtic.mil/docs/citations/AD1005582

Yeh, P. (2006). Using prediction markets to enhance US intelligence capabilities: A "Standard & Poors 500 index" for intelligence. *Studies in Intelligence* (50/4).

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net