Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Estimating Aggregate Data Center Investment with Project-level Data

Eirik Eylands Brandsaas, Daniel Garcia, Robert Kurtzman, Joseph Nichols, and Adelia Zytek

2025-109

Please cite this paper as:
Brandsaas, Eirik Eylands, Daniel Garcia, Robert Kurtzman, Joseph Nichols, and Adelia Zytek (2025). "Estimating Aggregate Data Center Investment with Project-level Data," Finance and Economics Discussion Series 2025-109. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2025.109.

Estimating Aggregate Data Center Investment with Project-level Data *

Eirik Eylands Brandsaas[†1], Daniel Garcia[‡1], Robert Kurtzman[§1], Joseph Nichols[¶1], and Adelia Zytek[‖1]

[1]Federal Reserve Board of Governors

December 17, 2025

Abstract

Data center investment in the U.S. has increased rapidly in the post-pandemic era, and plans for future investment have surged further. Forecasting investment at such a turning point is an important but potentially fraught exercise, especially given lags in aggregate data availability. We develop a straightforward method to forecast aggregate investment using project-level microdata and a small number of parameters: specifically, abandonment rates, time from plan-to-start, and time from start-to-completion. As a key validation of our approach, we generate estimates that match the recent history of aggregate data center investment in the NIPAs. We then use our method to generate nowcasts of aggregate data center investment in the short run, with the mean forecast indicating that investment will increase to $370 billion annualized by 2026:Q2. We can extend our methodology further out, but our forecasts then become conditional on the assumed flow of new data center plans. Assuming future plans range from one-fourth to twice the average pace of plans from 2024-2025 implies a range of investment forecasts of $360 billion to $930 billion in 2027, demonstrating the substantial upside and downside risks to future levels of investment.

Keywords: data centers, commercial real estate, forecasting, construction, time-to-plan

JEL Classification: R33, E22, E32, L74

*We thank Chris Kurz and Raven Molloy for their helpful feedback. The views expressed in this paper are solely those of the authors and do not necessarily reflect the opinions of the Federal Reserve Board or the Federal Reserve System.

†Division of Research & Statistics, Federal Reserve Board; eirik.e.brandsaas@frb.gov.
‡Division of Research & Statistics, Federal Reserve Board; daniel.i.garcia@frb.gov.
§Division of Research & Statistics, Federal Reserve Board; robert.j.kurtzman@frb.gov.
¶Division of Research & Statistics, Federal Reserve Board; joseph.b.nichols@frb.gov.
‖Division of Research & Statistics, Federal Reserve Board; adi.k.zytek@frb.gov.

1. INTRODUCTION

The post-pandemic era has seen a surge in data center investment, with investment in data center structures having roughly quadrupled from 2021 to 2025. Meanwhile, announced plans for future data centers have increased significantly, totaling over $1 trillion in the fall of 2025. Given these developments, an important question facing policymakers and practitioners is to understand the scale and dynamics of such investment in the short run and over longer horizons.

Our paper provides a methodology for answering this question that can generate aggregate real-time forecasts using microdata on project plans and a parsimonious number of empirically generated parameters. This approach provides an alternative framework for generating forecasts of non-stationary time series at the onset of an inflection point.[1] We apply our method to project-level data from the data center sector, which is in the midst of exactly such an inflection point and is of general interest given its role in the development of potentially transformative AI-powered services. Our method produces reasonable out-of-sample performance in the short run, and we show how it can be extended to longer horizons, highlighting the importance of assumptions on the flow of new plans in the future.

To generate a nowcast of future investment over the short run, defined as one-year out (the typical time from plan-to-start in our dataset), we require three parameters. For projects under construction, we require the time-to-completion. For projects in planning, we require the time from plan-to-start. Finally, we need the abandonment probability for the project, which is about 33 percent historically in the data center sector in our sample. To generate a forecast beyond one year, which we label as the medium run, one needs to make assumptions on the inflows of new plans. This is a more tenuous task, as it requires assumptions about the scope and value of future plans.[2]

We apply our methodology to estimate and nowcast data center investment over both the recent history and the short run using microdata on project plans from the Dodge Construction Network

[1]Our approach provides a forecast of the magnitude and timing of the inflection point before it is evident in the official statistics by using announced plans, instead of waiting until there are enough data after the inflection point to estimate a more traditional non-stationary time series model, such as regime switching or time-varying parameters models (Stock and Watson, 1996).

[2]To be precise as to how we are using our methodology over different time periods, we generate estimates of investment from the microdata over the period where we also have the official aggregate data. We generate nowcasts of investment over the period where we use only the currently reported project plans but do not yet have the official aggregate data. And, lastly, we generate forecasts of investment over longer periods that require assumptions of future plan activity.

(Dodge) and then extend the analysis by generating forecasts over the medium-term. An advantage of the Dodge data is that it is used as an input in the Census' Construction Put in Place (CPIP) survey, which informs BEA's investment measurements. Additionally, the project microdata typically includes total cost estimates, which allows us to speak to overall data center investment, as opposed to solely investment in the data center structure.

We first use the history of data center project plans reported in Dodge from 2003 to 2025 to generate estimates of the three key parameters discussed above. Applying these to the history of Dodge data center projects provides a quantitatively similar estimate to aggregate data center investment in the National Income and Product Accounts (NIPAs), providing a validation of our approach. We then use the Dodge microdata as of September 2025 to nowcast the near-term path of data center investment via a Monte Carlo simulation approach. In the mean of our simulations, investment increases from approximately $60 billion in 2024:Q4 to $180 billion in 2025:Q4, reaching an annual rate of roughly $370 billion by 2026:Q2. We then forecast investment through 2027 for three different scenarios of the future flow of new data center plans. These scenarios assume future plans range from one-fourth to double the average pace of plans from 2024-2025. In these simulations, mean annualized investment in 2027 ranges from about $360 billion in the more pessimistic alternative to $930 billion in a the more optimistic scenario; effectively, by the end of 2027, the level of investment is largely determined by the assumed inflow of new plans and not the current stock of plans.

As part of our work, we develop an approach for identifying data center specific investment in the NIPAs. Though aggregate data center investment in structures is already produced in the NIPAs, computer equipment and peripherals investment (E&I) is not broken out into data center investment. Because such E&I investment includes other product categories, such as business purchases of laptops and other computer equipment, we argue that one should remove the level of investment not attributed to data centers. We present a simple method to do this to construct a measure of aggregate data center investment in the NIPAs. Because the NIPAs come out with a lag, an additional advantage of our bottom-up approach is that it allows one to estimate such aggregate investment in real-time.

Our work provides an important complement to research on the broader economic impact of AI (see, e.g., Acemoglu 2025, Furman and Seamans 2019, or Brynjolfsson et al. 2025). While these papers are focused on the potential for AI to re-order the economy, primarily through increasing labor productivity, such potential developments are a function of the build-out of the underlying physical infrastructure of data centers. Our work thus provides a valuable complement to this literature with an innovative method that can translate detailed project level planning data into a time series of actual investment that is a pre-condition for the provision of next-generation AI services.

This analysis also contributes to the literature studying investment using project-level planning microdata. Glancy et al. (2025) show that abandonments out of planning are crucial determinants of short-run supply elasticities in the commercial construction space, providing motivation for our modeling approach. Millar et al. (2016) estimate time-to-plan for completed projects and highlight the significant heterogeneity in time from plan-to-start and start-to-completion across projects. Selgrad and Siani (2025) demonstrate the interaction of planned investment and monetary policy for firm-level projects, demonstrating the important role new projects play in responding to monetary policy shocks. Brandsaas et al. (2024) analyze plans for manufacturing activity, highlighting how starts rose around the time of the implementation of the Inflation Reduction Act and CHIPS and Science Act. Our work shows the value of plans in forecasting data center investment; the framework can naturally be applied to nonresidential structures investment, which typically averages about 20% of private investment (Brandsaas et al., 2024), or to other forms of investment where abandonments out of planning, time-to-plan, and time-to-construction can be consistently estimated.

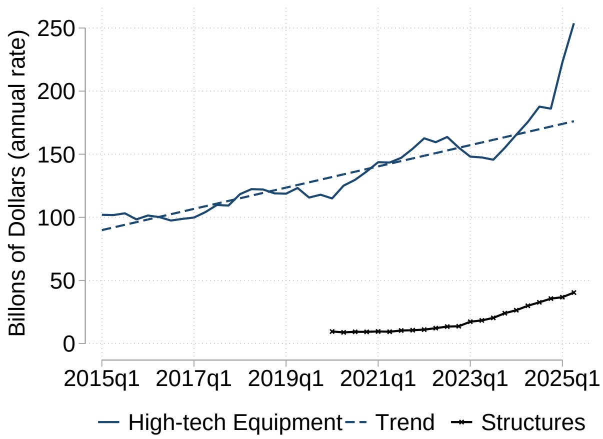
The rest of the paper follows as such. In Section 2, we describe how we obtain aggregate data center investment in the NIPAs. In Section 3, we present the framework we use for generating aggregate data center investment from project-level data. In Section 4, we describe the Dodge data and its properties. In Section 5, we present the simulation results. In Section 6, we conclude.

2. MEASUREMENT OF DATA CENTER INVESTMENT IN THE NIPAS

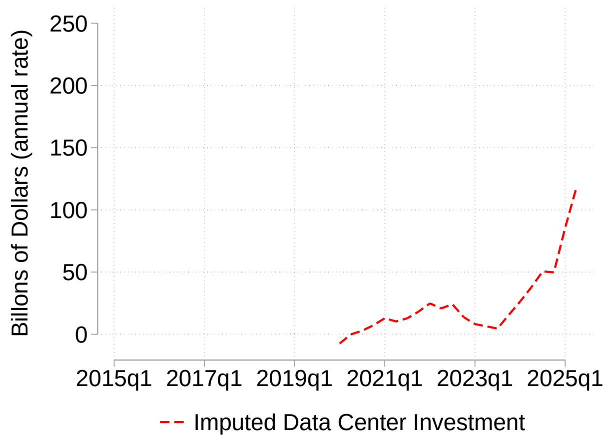
The NIPAs do not contain a single category of investment that captures all data center investment. Instead, investment related to data centers is recorded in a few main categories. BEA measures structure investment using the Census's nonresidential structures CPIP report. The left panel of Figure 1 shows that nominal spending on data center structures has increased rapidly, from about \$10 billion dollars in 2021 to about \$40 billion in the first half of 2025.[3]

However, total investment in data centers is much larger than the structures themselves, as data centers are equipped with high-tech capital goods, including servers, storage arrays, and networking

[3] Our investment simulations are in nominal terms, so here we plot nominal investment. Real investment in these categories is shown in Supplemental Materials Figure S1.



(a) Data Center Related Investment



(b) Imputed Data Center Investment

Figure 1: INVESTMENT IN DATA CENTER CATEGORIES. *Notes:* Panel (a): Nominal investment for data center structures and high-tech computers & peripheral equipment. The estimated trend is calculated from 2015 to 2022; see Section 2 for more details. Investment in data center structures is only available from 2020. Panel (b): Imputed data center investment, which is the sum of data center structures and detrended high-tech equipment.
Source: Authors' calculations using data from the Bureau of Economic Analysis.

hardware. The purchases of final computer hardware, including servers and other finished computing equipment, are counted in the "Computers and Peripheral Equipment" component of private fixed investment in information processing equipment, shown as the solid blue line in the left panel of Figure 1. The level of this series also reflects investment in other computer equipment, such as business purchases of laptops and printers.

Our approach is to effectively detrend such computer E&I investment and take the residual as the E&I investment in data centers. We estimate a time trend from 2015 to 2022, shown as the dotted line in the left panel.[4] This is an estimate of counterfactual investment absent the recent surge in data center activity. Next, we interpret the difference between the solid and dotted lines as reflecting data center equipment investment, and we add this difference to structures investment (black line, left panel) to obtain a measure of data center investment.

This sum is shown in the right panel of Figure 1. Though we think this measure of data center

[4]We estimate the trend through 2022, to predate the start of the ramp up in data centers structures spending in 2023. That said, the trends are similar when estimated through 2021 or 2023.

investment accurately captures recent trends in data center investment, we note this is still an imperfect estimate. On the one hand, it does not include some relevant pieces, such as investment in specialized data center software or own-account equipment investment.[5] On the other hand, this estimate could include some purchases of equipment unrelated to data centers, if, say, business purchases of laptops increased meaningfully above trend.

3. A FRAMEWORK FOR FORECASTING AGGREGATE INVESTMENT WITH PROJECT-LEVEL PLANNING DATA

In this section, we describe the framework we use to simulate aggregate data center investment using project-level data.

3.1. Environment

Our analysis is indexed at the project level, and each project i has a vector of characteristics, X_i. These characteristics include an estimate of total project cost in dollars V_i. The project evolves through different stages, j. The project stages can take one of four values: in planning (P), under construction or "started" (S), completed (C), or abandoned (A).

All projects are assumed to start with a planning stage (P), and then after some number of periods advance to a construction start or are abandoned. Once started, projects are assumed to reach completion after some additional number of periods.[6]

The average time from plan to construction, $T_{P,S}(X_i)$, can vary with project characteristics, in particular the size of the project. The average time from start-to-completion, $T_{S,C}(X_i)$, can similarly vary with project characteristics. The abandonment rate, $\lambda(X_i)$, is the average number of projects that are abandoned out of planning at the onset of our simulation, which can also vary with project characteristics.

There are various potential assumptions one could make with regards to the distribution of investment across the construction stage—that is, how quickly spending is phased in once construction begins.

[5] We do not include software investment because recent changes in total software investment likely reflect other AI developments beyond investments in software to operate data centers themselves. For more context on own-account equipment investment, see Byrne et al. (2017) and Byrne et al. (2018). The in-house assemblies of intermediate capital purchases are not included in the equipment series above.

[6] A small share of projects that reach construction are abandoned—we can observe and account for these cases once they occur—but for estimating to plan or time-to-completion, given their small share, we assume this case away.

For our analysis, we assume that investment is distributed evenly across the construction timeline. In effect, a project with value V_i and months from start-to-completion of $T_{S,C}$, spends out $I_i = V_i/T_{S,C}(X_i)$ each month from its start date through completion. This assumption would be too backloaded if we were only considering building construction. The U.S. Census Bureau estimates that for very large projects ($>$ \$100 million) that take over 3 years to fully build out, typically about 50 percent of the value is phased-in the first year, and 90 percent by the end of the second year.[7] However, it seems reasonable to assume that equipment installation will be relatively more backloaded than construction of the structure itself. Therefore, this assumption balances these competing forces; nonetheless, this assumption can be easily varied within our method.

In the simulations and discussion, we find it helpful to distinguish between the "short run" and the "medium run." In the short run—about a year from the last vintage of planning data—the level of investment is largely determined by the stock of current, existing plans, since it typically takes about one year for plans to move from planning to the constructions start stage. While investment in the short run is largely a factor of existing plans, investment in the medium run is instead heavily dependent on assumptions about future inflow of plans.

3.2. Nowcasting the Short Run

Given the three parameters defined above and a stock of projects still in the planning stage or which are still under construction at time t, we can generate a nowcast of the aggregate time series for the value-in-place from these projects. The first step is to denote the month a project starts construction as s_i. For projects that have observed starts in the data, we know the start date, while for projects that are plans (that have not been cancelled in the simulation), s_i can be derived from the month the project entered the data and the time from plan-to-start $T_{P,S}(X_i)$.

The total value of investment value-in-place in month t is given by summing across all projects in construction in month t:

$$\mathscr{I}_t = \sum_i \mathbb{1}[s_i \leq t < s_i + T_{S,C}(X_i)] \cdot I_i, \tag{1}$$

where $\mathbb{1}[\cdot]$ is an indicator function equal to 1 if project i is under construction in month t and 0 otherwise, and $I_i = V_i/n$ is monthly project-level investment defined as above. Thus, a project that is under construction for, say, twelve months ($T_{S,C}(X_i) = 12$) and started construction in January 2025 will contribute to investment through the entire year. We note that this framework for investment—

[7] For more information, see https://www.census.gov/construction/c30/pdf/t424.pdf.

of spending that is phased-in while the project is under construction—is closely related to the methodology employed by U.S. Census Bureau to estimate investment on nonresidential structures. However, this process does differ from how the contribution from equipment is measured, which the U.S. Census Bureau measures at the time of purchase.

3.3. Forecasting the Medium Run

In order to extend our simulations beyond one-year ahead, assumptions are required about the the inflow of new projects into planning, unlike the short-term nowcast which depends on the actual announced stock of project plans.

Let $\bar{t}$ denote the last month for which we observe planned projects, and let $\overline{T_{P,S}}$ denote the average time from planning to construction start across all characteristics. New projects entering planning after $\bar{t}$ will not meaningfully contribute to investment until month $\bar{t} + \overline{T_{P,S}}$. We define the start of the medium run as:

$$\bar{k} = \bar{t} + \overline{T_{P,S}} \tag{2}$$

For example, if $\bar{t}$ is December 2025 and $\overline{T_{P,S}}$ is 10 months, then $\bar{k}$ is September 2026. In practice, investment in month $\bar{k}$ will be dominated by projects from the existing stock with start dates $s_i < \bar{k}$. As the forecast period increases, the share of investment attributable to new project inflows increases.

4. PROJECT-LEVEL DATA

This section first describes the project-level data we use to generate aggregate measures of data center investment. We then present descriptive statistics from our data sample. Last, we provide detail on the sample used in our simulation exercise.

4.1. Dodge Construction Network Data on Data Centers

Our analysis is based on Real Estate Analyzer (REA) data from Dodge Construction Network from 2003 to September 2025. Dodge is a leading data provider in the construction industry and their data are used by the U.S. Census Bureau to measure CPIP. Specifically, the Census surveys a stratified sample of construction projects obtained from Dodge, and project managers then report on the value of work ("value-in-place") done each month from project start-to-completion. Previous work, for instance Brandsaas et al. (2023), has shown that data on Dodge construction starts are helpful for predicting future spending on structures.

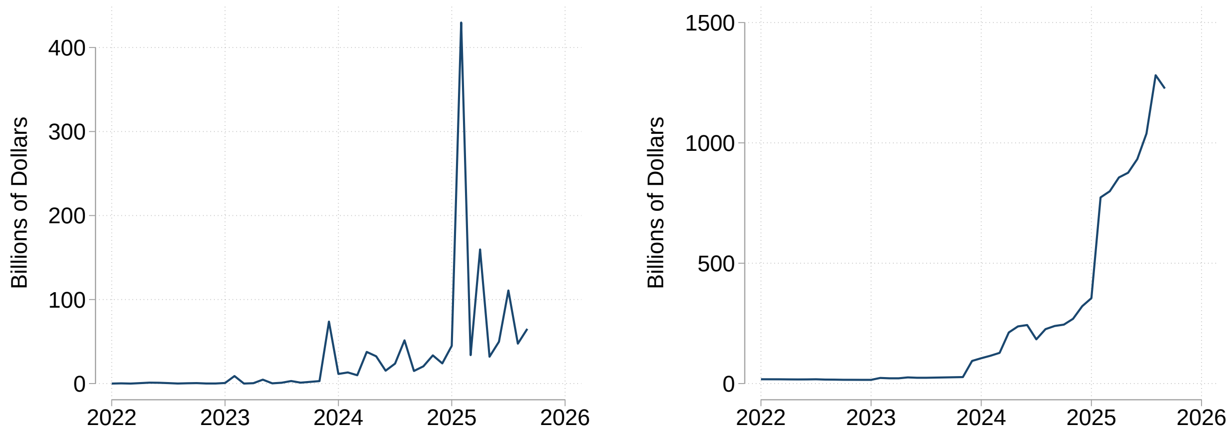
In our work here, we instead mainly focus on construction project *plans* rather than starts for a few key reasons. First, plans increase the lead time for economic projections, because they lead construction starts by about a year on average. Second, in the REA data, the dollar value of project-level plans likely mainly reflects total costs for the project (including both equipment and structure). We verify this is the case by comparing plan costs in REA versus costs we found online for the largest recently announced data center projects. The project-level plans are especially valuable in this context since we care about total investment rather than structures investment alone.[8] Basing our approach on actual project level plans is thus useful for estimating and nowcasting overall investment in data centers related to both equipment and structure as measured in the NIPAs (as described in Section 2).[9]

The REA data contain information on each project's property type, square footage, location, and project value. The total project values we use are in nominal US dollars and we only consider projects being built in the United States. We take the last reported plan value to be the value of the plan. We consolidate project stages into four categories: planning, start, completed, and abandoned. The analysis requires harmonizing the REA data in order to consistently track projects from planning to subsequent phases. For instance, we assess a project to have entered the start stage the last time it does so without becoming a plan again. We also assume abandons and completions can only occur at the end of a project's series. All possible phase changes are shown in Figure S4 and information on how these phase changes are derived can be found in Section S.2.1, both in the Supplementary Materials. Some large projects are initially reported as master reports and then receive child reports for individual structures or building phases.[10] Dodge provides a mapping between master and child

[8] In the REA data, project-level values can be updated over time, and in our simulations we use the last available estimate of project value (as of September 2025). Given Dodge's focus on the construction industry, it is possible that some project-level costs revise down to hone in on structure costs, as the project moves from planning to construction stage. While project values do revise, they do not show a strong tendency to revise down (or up) as they move through construction stages. Hence, our interpretation that project values mainly reflect total project costs.

[9] Industry analysts have imputed the level of data center investment using other approaches. For instance, analysts at J.P. Morgan (Reinhart and Feroli, 2025) forecast investment based on estimates of announced data center power consumption and assumptions on the costs of achieving those planned goals. Analysts at Goldman Sachs (Peng et al., 2025) generate estimates using changes in revenue or reported investment in companies or industries exposed to the AI boom.

[10] To focus on data centers, we subset the data to all projects whose most recently recorded structure type is a data center, as well as the master reports for these projects regardless of their structure type.



(a) Value of New Plans over Time

(b) Planning Stock Value over Time

Figure 2: VALUE OF NEW PLANS AND PLANNING STOCK OVER TIME. *Notes:* Panel (a) shows the value of new data center plans in a given month. Panel (b) shows the total stock of data center projects currently in the planning stage in a given month.
Source: Authors' calculations using data from Dodge Construction Network.

reports; our process for linking these reports is outlined in the Supplementary Materials in Section S.2.2.

4.2. Descriptive Statistics

The left panel of Figure 2 plots monthly unadjusted nominal value of new data center plans in the Dodge data.[11] Prior to 2023, data center plans were modest in scale; for instance, they averaged about $7 billion per year from 2019 to 2022. However, data center plans began to skyrocket in late 2023, cumulating to new plan values of almost $300 billion in 2024 and over $900 billion in 2025 (as of September). As some of these plans entered the construction phase, they contributed to greater data center spending, as shown in Figure 1. However, most of these plans have not yet broken ground. The right panel of Figure 2 shows that the value of the stock of data center plans—which cumulates the inflow of plans while excluding plans that have begun construction or have been abandoned—exceeds over $1 trillion dollars as of September 2025.

The increase in plans reflects both an increase in the number of announced data center projects, as

[11] We choose to present nominal spending (rather than real spending), as forecasting deflators for spending on data center structures and equipment would introduce another layer of uncertainty.

well as an increase in the dollar value of the plans. For instance, in the four years from 2019 to 2022, the data includes plans for about 278 data centers, with an average project value of about \$114 million and median value of \$26 million. In 2025 through September, there were 521 plans, with an average value of a little under \$2 billion and median value of about \$216 million.

4.3. Estimation sample

As discussed in Section 3, to simulate investment we require estimates of time from plan-to-start, time from start-to-completion, and abandonment rates. To obtain these estimates, we focus on a sample of projects with observed planning stages and with observed transitions out of planning.[12] This drops many recent plans from the estimation sample (since it is too early to determine whether they will build out or not), but we include these later in the simulation sample. We also exclude master reports, which aggregate multiple projects and have ambiguous abandonment status, and these projects are also reintroduced in the simulation sample. In total, these reductions reduce the estimation sample size by 56%, predominantly due to recent projects that have not transitioned out of planning.

Table 1 below shows summary statistics for abandonment rates, time in months from plan-to-start and start-to-completion, and project values for all projects in the estimation sample in panel A, and for relatively large projects—those worth over \$250 million dollars, corresponding roughly to the top decile of plans—in panel B. Starting with panel A, which is based on about 1,700 projects, we estimate an abandonment rate of 0.33. Meanwhile, it takes about 11 months for projects to move from plan-to-start, and on average, about 9 months for projects to move from start-to-completion, as recorded in the REA data. The statistics on months of time from plan-to-start, and from start-to-completion, are similar to those reported in Glancy et al. (2025) for the overall commercial sector. However, the abandonment rate is lower than the 0.46 abandonment rate reported in Glancy et al. (2025), suggesting data centers have somewhat lower abandonment rates, typically, than typical office or retail buildings.

Turning to panel B, for the relatively larger projects the abandonment rate is about 0.3, similar to the rest of the sample. Consistent with Glancy et al. (2025), we find that abandonment rates do not vary much (unconditionally) by size, perhaps reflecting countervailing forces. On the

[12]We classify projects as abandoned once they have been observed in planning for at least 48 without an observed start phase, since about 95% of projects that begin construction do so within four years, see Figure S2 in the Supplemental Materials.

	Mean	S.D.	p10	p50	p90	Count
Panel A: All projects						
Fraction abandoned (λ)	0.33	0.47	0.00	0.00	1.00	1711
Months plan to start ($T_{P,S}$)	11.11	15.36	2.00	6.00	26.00	981
Months start to compl ($T_{S,C}$)	9.37	6.83	3.00	9.00	17.00	715
Value in billions (V)	0.10	0.38	0.00	0.01	0.25	1711
Panel B: Large plans (over 250 million USD)						
Fraction abandoned (λ)	0.30	0.46	0.00	0.00	1.00	169
Months plan to start ($T_{P,S}$)	15.00	18.47	2.00	9.00	37.00	113
Months start to compl ($T_{S,C}$)	19.77	9.78	10.00	18.00	35.00	48
Value in billions (V)	0.70	1.02	0.28	0.48	1.07	169

Table 1: ESTIMATION SAMPLE SUMMARY STATISTICS. *Notes:* The table shows summary statistics (mean, standard deviation, percentiles of the distribution, and count) for data center projects in the estimation sample. Months from plan-to-start and start-to-completion (compl) are only calculated for projects with start and completion entries, respectively.
Source: Authors' calculations using data from Dodge Construction Network.

one hand, abandonment rates for large projects may be more sensitive to changes in economic conditions, as suggested by Glancy et al. (2025). On the other hand, the firms involved in larger projects are likely to differ—for example, they potentially have easier access to credit. Turning to construction timelines, we find months from plan-to-start of about 15 months for these larger projects, a little higher than the 11 months in the overall sample (not statistically different). There is a more meaningful difference in months from start-to-completion (20 vs 9 months), showing that, as expected, larger projects take longer to construct.

5. DATA CENTER INVESTMENT SIMULATION

In this section, we present the results from our simulation exercise. We first present the calibration details; we then present results for the short run and medium run, with the latter requiring assumptions about the inflow of new plans going forward.

5.1. *Calibration*

We calibrate the parameter distributions based on the estimation sample as follows.

For the distribution of the abandonment rate λ, we set the mean to 0.33 and the standard deviation to the standard error of the mean 0.01. For the duration of months from plan-to-start $T_{P,S}$, we set the mean to 11.11 and the standard deviation to 0.49.[13]

Because time from start-to-completion $T_{S,C}(X_i)$ has a strong association with project size, we model it as follows:

$$T_{S,C}(X_i) = \beta_0 + \beta_1 \cdot \log(\mathrm{V}_i). \tag{3}$$

This regression yields $\hat{\beta}_0 = 19.15$ (SE = 0.54) and $\hat{\beta}_1 = 2.15$ (SE = 0.11). This equation implies a \$1 billion data center project typically takes about 19 months from start-to-completion.

The discussion above describes our treatment of plans. For projects that we see in the data that have already begun construction, we simulate spending based on the observed start date and the time-to-completion calibration approach described above.

5.2. *Short-Run Nowcasts*

The sample used in the simulation includes all projects that are in planning or that have started but are not completed or abandoned in the dataset provided by Dodge in September 2025. In the simulations, there are two broad sources of randomness. First, in each simulation key parameters are independently drawn from normal distributions, as described in the previous subsection. Second, in each simulation, abandonments are determined stochastically at the project level. In some simulations, many of the very large projects end up abandoned, and, as a result, total investment is relatively lower, and vice versa. In practice, most of the randomness comes from the project-specific abandonment.

To generate nowcasts, we use the method described in Section 3. We generate 1000 Monte Carlo simulations of investment over the short run. After drawing project-level abandonments, we obtain a set of non-abandoned projects. Investment is then simulated for these projects based on the time-to-build parameters and the assumptions described above.

The left panel of Figure 3 presents the simulation results for the short-run period through 2026:Q2.

[13] The means are reported in the first column of Table 1. The standard error of the mean is given by the ratio of the standard deviation (second column) and the square root of the number of observations (last column).

The figure displays data center investment at a quarterly frequency (annualized) across various distribution points: mean, median, interquartile range, and additional percentiles. In the mean simulation, investment increases from approximately $60 billion in 2024:Q4 to $180 billion in 2025:Q4, reaching an annual rate of roughly $370 billion by 2026:Q2.

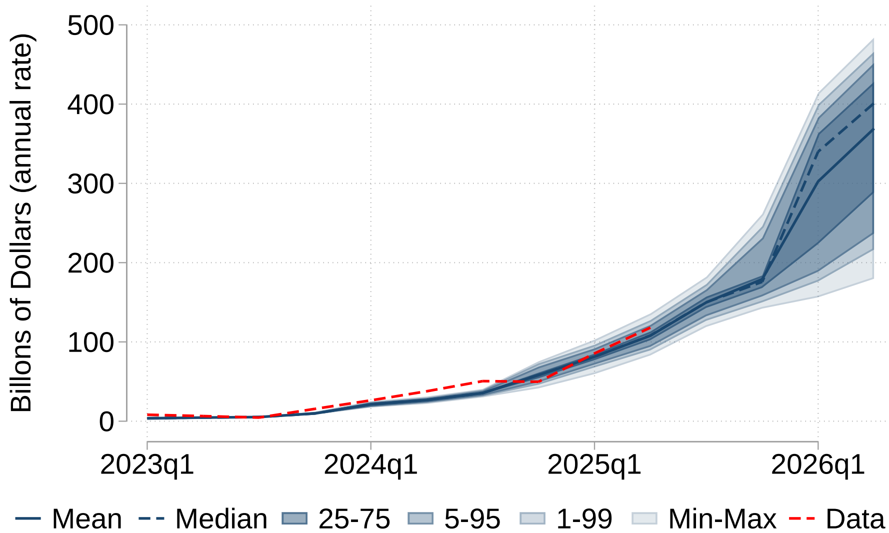
As a validation exercise, we compare simulated nominal investment to the NIPA-based imputation of data center investment discussed in Section 2. This estimate is shown as the red dotted line in the left panel of Figure 3, and is the sum of nominal data center structures with “excess” high-tech equipment, defined as the difference between measured investment in the computers and peripheral equipment category and a trend estimated from 2015 to 2022. In the figure, the red line is similar to simulated investment, indicating that, thus far, simulated investment is a good guide for actual investment.

The simulations exhibit substantial dispersion. At the 5th percentile, investment reaches approximately $240 billion in 2026:Q2, while investment stands at about $450 billion at the 95th percentile. This variation primarily stems from project-level abandonment rates.[14] In some simulations, many recently announced large data centers are completed, while in others, many abandons occur.

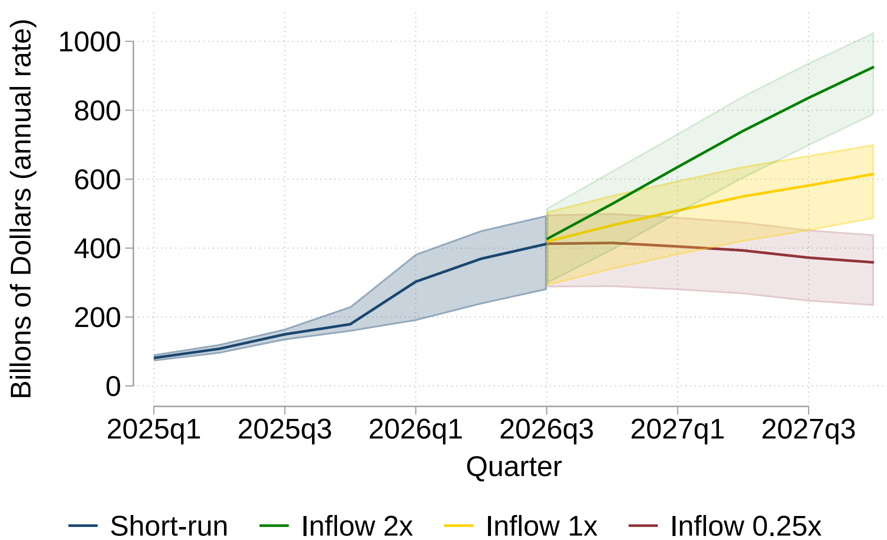
The median exceeds the mean in these simulations because the distribution of total investment, shown in Figure S3 in the Supplemental Materials, displays bimodality with greater mass in the right mode. This distribution pattern results from two factors: the abandonment rate of 0.33, and the right-skewed distribution of project-level investment values. Specifically, a small number of extremely large projects, such as Project Stargate, represent significant outliers. Simulations where these outsized projects are abandoned yield lower overall investment levels. Conversely, when completed, they substantially elevate investment totals. Since projects survive approximately two-thirds of the time in the model, the right-side distribution mode contains greater mass. Hence, most simulations include the completion of the very large projects, thereby pushing the median above the mean.

We interpret this nowcast as consistent with the Gabaix (2011) “granular origins” view of aggregate fluctuations. The distribution of data center projects has a fat right tail, so the magnitude of the investment boom depends critically on the spending and completion of a handful of very large

[14] The confidence intervals start prior to the latest vintage of Dodge data, because in the simulations, projects can break ground before they enter the construction stage in the Dodge data. As an example, a project that enters the planning stage in 2024:Q1 could break ground in some simulations later in the year even if the Dodge data do not yet indicate the project has begun.



(a) Short Run



(b) Medium Run

Figure 3: DATA CENTER INVESTMENT. *Notes:* Panel (a) displays simulated short-run investment based on 1000 simulations. The red dotted line represents data center investment imputed from NIPA data (see Section 2). In panel (b), we simulate through 2027 based on three different scenarios for the level of new plans in the future based on the lagged-level of new plans times a scaling factor (2 in green, 1 in yellow, or 0.25 in red). The shaded areas denote the 5th-95th percentiles.
Source: Authors' calculations using data from Dodge Construction Network.

investment projects. That said, it is worth noting that the rise in data center projects has been sufficiently broad-based that, even in the most pessimistic scenarios where many of these large projects are completely abandoned, investment still rises to $240 billion by the middle of 2026.

5.3. Medium-Run Forecasts

The simulations through the short run are independent of our assumptions about the future inflow of plans, as plans take about a year in the simulations before they break ground. We now turn to simulating investment through 2027 (which we label as the medium run), which requires assumptions about the inflow of new plans through 2027. We first calculate the average plan size and the monthly flow of new plans from 2024 to September 2025. We then consider three alternatives for the number of plans going forward to provide a range of scenarios: a) one-fourth, b) the same, or c) twice the number of plans per month in this recent history, holding the project size constant. Lastly, we simulate investment through 2027 according to these three different scenarios; the results are shown in the bottom panel of Figure 3 and the first three columns of Table 2. The figure includes the 5th and 95th percentiles for each simulation.

The simulations start to branch out in late 2026, when the assumed plans in late 2025 start breaking ground. By the end of 2027, investment differs markedly depending on the assumptions about plans, with mean investment ranging from about $360 billion in the more pessimistic alternative to $930 billion in the more optimistic scenario. Investment is relatively robust, even in the scenario where future plans fall to a quarter of their recent level, as data center projects currently in planning support investment for a few years. Hence, the 2027 forecasts reflect a mix of both current "data"—plans that have already been announced—as well as the simulated paths of plans going forward. Looking beyond 2027, simulated investment would rely even more heavily on the assumptions about new plans.

Table 2 also reports the data center investment contributions to nominal GDP growth from the simulations. To calculate the contributions, we grow out nominal GDP in 2024 using nominal GDP growth forecasts available from the Survey of Professional Forecasters (SPF).[15] In the middle three columns, we report these contributions before accounting for the share of high-tech equipment that is imported. In the mean simulation, data center investment contributes 0.4 percentage points to

[15]The SPF does not include a 2027 nominal GDP forecast. To obtain it, we use the real GDP growth forecast and an extrapolation of the GDP price index from previous values. Since these forecasts likely incorporate some boost from data center growth, we do not adjust nominal GDP in the three different scenarios.

	Investment (Bill. $)			Contributions to GDP Growth (p.p.)					
				Including Imported Inputs			Excluding Imported Inputs		
Year	0.25x	1x	2x	0.25x	1x	2x	0.25x	1x	2x
2023	10	10	10	0.02	0.02	0.02	0.01	0.01	0.01
2024	58	58	58	0.17	0.17	0.17	0.07	0.07	0.07
2025	179	179	179	0.41	0.41	0.41	0.18	0.18	0.18
2026	413	464	526	0.75	0.91	1.11	0.33	0.40	0.49
2027	357	612	921	-0.17	0.45	1.21	-0.08	0.20	0.53

Table 2: DATA CENTER INVESTMENT & CONTRIBUTION TO GDP GROWTH BY INFLOW OF FUTURE PLANS.
Notes: The table shows mean data center investment in the fourth quarter (annual rate) and investment contributions to GDP growth based on different assumptions about the future inflow of plans as described in Section 5.3. The investment contribution is calculated as the share of GDP in the previous year times the Q4/Q4 growth rate of data center investment. Our import adjustment assumes that 44% of investment is domestic; see Section 5.3 for details. Nominal GDP growth comes from the Survey of Professional Forecasters (SPF). Because the SPF does not include a 2027 nominal GDP forecast, we use the real GDP growth forecast and an extrapolation of the GDP price index from previous values. SPF data is available through the Research Department, Federal Reserve Bank of Philadelphia, at https://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters.
Source: Authors' calculations using data from Dodge Construction Network and the SPF.

GDP growth in 2025. In 2026, the contributions range from about 0.8 to 1.1 percentage points, depending on the assumptions about future plans. Finally, in 2027, the contributions to GDP growth differ markedly, ranging from 1.2 percentage points in the more optimistic scenario to -0.2 percentage points in the more pessimistic one.

However, the net effect of the data center investment boom on GDP growth will not only depend on the level of investment, the main focus of our analysis, but on the share of investment net of imports. In turn, we require an estimate of the import share of data center investment; we create a "ballpark" estimate in the following way. First, we assume that 30 percent of investment is in domestic structures and the remaining 70 percent in high-tech equipment. We derive this split using BEA and Census estimates of data center structure value put in place and high-tech and other equipment spending by information and data processing firms in 2024. Second, we assume an import share of 80% for high tech equipment. The import share is based on imports and total investment in high tech equipment from BEA and Census, also in 2024. Combining these assumptions implies that 44% of data center investment is domestic. The resulting contributions are shown in the final three columns of Table 2, and are just 44% of the middle three columns.

Importantly, the import-adjustment adds another level of imprecision to our forecasts. An advantage of our method in the short run is that it generates nowcasts for the total level of data center investment (and hence total GDP contributions) with very little judgement, as it is largely pinned down by the current stock of plans and historical estimates for the model parameters. However, accounting for imports requires more judgement about the import shares in coming quarters. The medium run forecasts for data centers' contribution to GDP growth require assumptions on both import shares and the inflow of plans; given that both margins are currently unknown, there is substantial uncertainty around any forecast of such investment in the medium run.

5.4. Further Robustness and Discussion

We briefly discuss how changes to key parameters would affect the simulations. The abandonment rate of 0.33 was chosen from the historical average, but the yearly abandonment rates vary from about 0.2 to 0.45. Changes in the abandonment rate can have subtle effects on the percentiles of the distribution of simulated investment. For example, if abandonment rates rise above 0.5, the distribution of investment would remain bimodal, but with greater mass in the left mode, triggering a downward shift in the median. However, the mean scales roughly proportionally with abandonment rates. For instance, a change in the abandonment rate from 0.33 to 0.45 would roughly imply scaling down mean investment levels by 18 percent (as the survival rate shifts down 18 percent from 67 to 55 percent). One might expect the abandonment rate parameter to increase if the expectations around the potential returns to AI were to worsen. One advantage of the bottom-up simulation approach developed in this paper is that it can dynamically adjust as such abandonments occur.

A related discussion pertains to scaling factors over the lifetime of the project; that is, it is possible that many of the projects survive and build out, but are eventually scaled down or up in scope. In the historical data center data, we did not see a strong tendency for values to scale one way or another, and so we opted to leave scaling factors out of the baseline. With a different assumption, mean investment would scale proportionately, so long as the scaling factor applies equally to all projects.

It also seems plausible that the unprecedented acceleration in data center construction could be delayed by production capacity constraints. While predicting when and where supply chain bottlenecks manifest is notoriously challenging, there are signs of strain across many of the inputs needed to build data centers, including land, labor, equipment, power, and capital (see, e.g., Chen (2025), Turner and West (2025), Duguid and Kinder (2025), Whelan (2025)). The simulations could be adjusted with different assumptions about time from plan-to-start or start-to-completion. Shifts in the construction timeline would roughly delay simulated investment proportionally.

6. CONCLUSION

We provide a new method to generate real-time estimates and short-run nowcasts of aggregate spending on data center investment based on project-level data using a small number of empirically estimated parameters. This method can also generate medium-run forecasts with assumptions on the inflow of future plans. Our estimates performs well out-of-sample into 2025, and so we use it to nowcast investment in 2026 given observed planning and construction activity as of September 2025. We show how to use our methodology to forecast further out as well, but these medium-run forecasts become quite sensitive to assumptions about inflows of planning activity which are currently unknown.

Though data center investment is of particular interest given the large investments in the space during the post-pandemic era, our methodology can be valuable for examining investment of any category during potential inflection points, as long as comprehensive project-level microdata are available. That said, our simulations do not take into account the vast array of other macroeconomic effects from the surge in data center investment or the adoption of artificial intelligence. Separating out these channels as the economy evolves will be an important challenge for future research.

References

Acemoglu, Daron, "The Simple Macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58.

Brandsaas, Eirik E., Daniel I. García, Joseph B. Nichols, and Kyra Sadovi, "Nonresidential Construction Spending is Likely Not as Weak as it Seems," *FEDS Notes. Washington: Board of Governors of the Federal Reserve System*, 2023.

Brandsaas, Eirik E., Robert J. Kurtzman, and Joseph B. Nichols, "From Plans to Starts: Examining Recent Trends in Manufacturing Plant Construction," *FEDS Notes. Washington: Board of Governors of the Federal Reserve System*, 2024.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond, "Generative AI at Work," *The Quarterly Journal of Economics*, May 2025, *140*, 889–942.

Byrne, David M., Carol Corrado, and Daniel E. Sichel, "Own-Account IT Equipment Investment," *FEDS Notes. Washington: Board of Governors of the Federal Reserve System*, 2017.

Byrne, David M., Carol Corrado, and Daniel E. Sichel, "The Rise of Cloud Computing: Minding Your P's, Q's and K's," *NBER Working Paper No. 25188*, 2018.

Chen, Te-Ping, "Data Centers are a 'Gold Rush' for Construction Workers," *Wall Street Journal*, https://www.wsj.com/business/data-centers-are-a-gold-rush-for-construction-workers-6e3c5ce0. November 29 2025.

Duguid, Kate and Tabby Kinder, "Investor Angst Over Big Tech's AI Spending Spills Over into Bond Market," *Financial Times*, https://www.ft.com/content/d2bf6c25-fb42-4f13-b81c-a72883632f50. November 11 2025.

Furman, Jason and Robert Seamans, "AI and the Economy," *Innovation Policy and the Economy*, January 2019, *19*, 161–191.

Gabaix, Xavier, "The Granular Origins of Aggregate Fluctuations," *Econometrica*, May 2011, *79* (3), 733–772.

Glancy, David, Robert J. Kurtzman, and Lara Loewenstein, "JUE Insight: Shovel Ready Projects and Commercial Construction Activity's Long and Variable Lags," *Journal of Urban Economics*, 2025.

Millar, Jonathan N., Stephen D. Oliner, and Daniel E. Sichel, "Time-to-Plan Lags for Commercial Construction Projects," *Regional Science and Urban Economics*, 2016, *59*, 75–89.

Peng, Elsie, Joseph Briggs, and Sarah Dong, "US Daily: A Primer on the Impact of AI on the GDP Statistics," *Goldman Sachs Economic Research*, September 13 2025.

Reinhart, Abiel and Micheal Feroli, "US: The Data on Data Centers," *J.P. Morgan: North America Economic Research*, January 15 2025.

Selgrad, Julia and Kerry Siani, "Monetary Policy and Investment Plans," *Working Paper*, 2025.

Stock, James H. and Mark W. Watson, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 1996, *14* (1), 11–30.

Turner, Nicol and Darrell West, "The Future of Data Centers," *Brookings*, https://www.brookings.edu/articles/the-future-of-data-centers/. November 5 2025.

Whelan, Robbie, "CoreWeave Reports Doubling of Revenue From AI Boom," *Wall Street Journal*, https://www.wsj.com/business/earnings/coreweave-earnings-q3-2025-crwv-stock-cd0f9c8c?. November 10, 2025.

SUPPLEMENTARY MATERIALS

This document contains the supplementary materials as referenced in the manuscript. The first section contains the supplementary figures that are referenced in the text. The second section provides further detail on how we clean and process the data from Dodge Construction Network (Dodge).

S.1. Supplementary Figures Referenced in Text

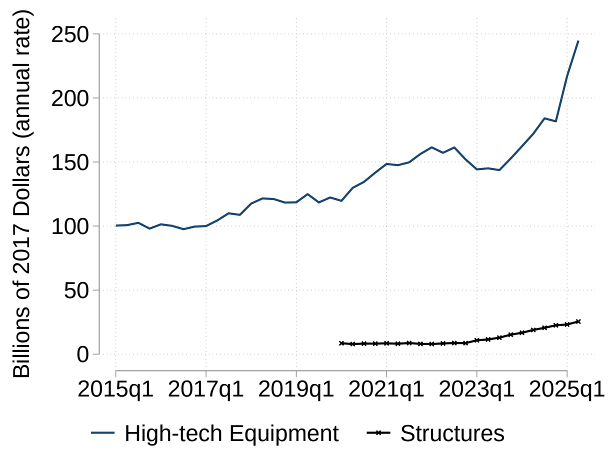


Figure S1: INVESTMENT IN DATA CENTER CATEGORIES. *Notes:* Real investment by quarter for data center structures and high-tech computers & peripheral equipment.
Source: Authors' calculations using data from the Bureau of Economic Analysis.

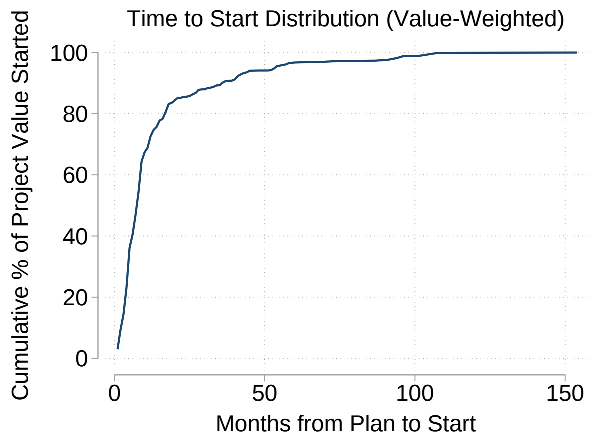
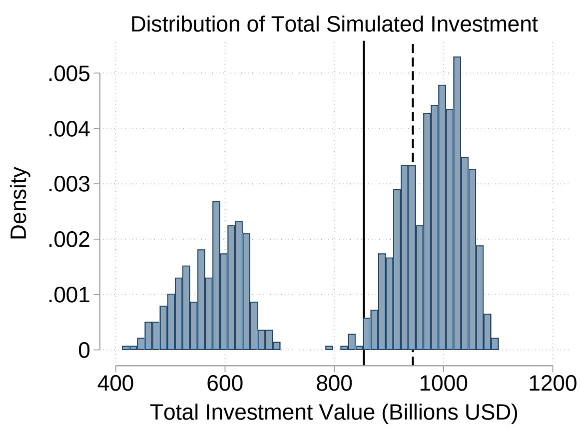
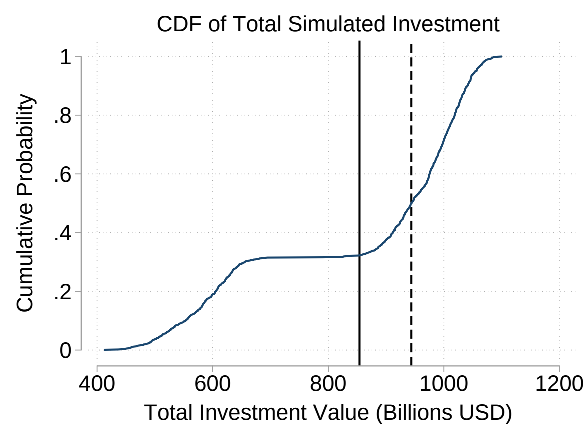


Figure S2: TIME-TO-START CDF. *Notes:* Distribution of months from plan-to-start across all projects with observed plan and start phases.
Source: Authors' calculations using data from Dodge Construction Network.



(a) Density



(b) CDF

Figure S3: DISTRIBUTION OF SHORT-RUN SIMULATED INVESTMENT VOLUME. *Notes:* Values are based on 1000 simulations of short-run investment. The vertical black lines denote the mean (solid) and median (dashed).
Source: Authors' calculations using data from Dodge Construction Network.

S.2. Data Construction Details

This section discusses how we harmonize the project-level Dodge data so that construction phases are defined consistently and transitions between phases conform to a consistent structure. This section also discusses how we integrate master reports and their linked child reports, as well as other data adjustments.

S.2.1. Phase Work

In the original Dodge reports, projects can move in and out of any of the eight phases designated by Dodge: pre-planning, planning, final planning, bidding, underway, completed, abandoned, and deferred. We make adjustments to create a consistent phase structure across projects. First, we combine pre-planning, planning, final planning, and bidding into a single ”planning” phase. We then make phase adjustments such that only phase changes shown in Figure S4 are permissible. We achieve this structure by making the following adjustments to the data in sequential order:

1. For each project, the last transition into the **start** phase is found. All preceding entries have their phase changed to **plan**.
2. For each project, the last transition into the **plan** phase is found. All preceding entries have their phase changed to **plan**.
3. For each project, any series of **abandons/deferrals** that do not come at the end of the project are switched to either **plan** or **start**, depending on which of those two phases most recently preceded the string of abandons/deferrals. Additionally, any project that is in the **planning** stage for 48 months and their last stage label is abandoned or plan, has their phase changed to abandoned, starting at the 48+1 month mark. This is because about 94% of projects (weighted by project value) with observed start dates start within 48 months (see Figure S2).
4. **Completions** can only come at the end of the series of reports because of how the completion variable is defined: every entry after the last updated completion date is considered a completion.

Taken together, these changes ensure the following:

1. **Abandoned or completed** entries can only occur at the end of the series of entries for a project. A project can no longer be labeled deferred. If a project is ongoing and is currently deferred, it is labeled as abandoned.

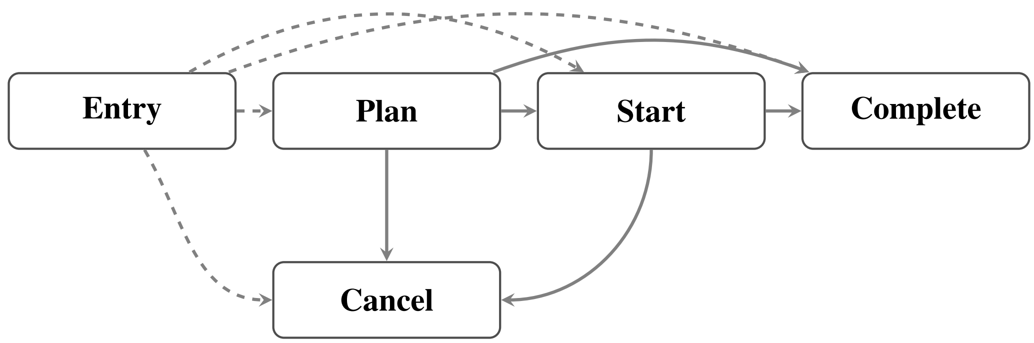


Figure S4: VALID PHASE TRANSITIONS. *Note:* Solid arrows indicate allowable transitions. The dashed arrows from entry represents the conceptual point at which a project enters the database, rather than an observed phase.

2. Once a project moves out of a **plan**, it will never become a plan again. Once a project moves out of a **start** it will never move back into a start (see arrows in Figure S4).
3. Projects can enter the data set at any of the four redefined phases.

S.2.2. Merging Master Reports with Child Reports

Larger projects are often first reported as master reports and, when additional data are available, child reports are made for specific components or phases of the project. Child reports receive unique project identifiers and Dodge provides a crosswalk linking master reports with their child report, once the child report has received its first entry.

We merge child reports with their master report in the following way. For all entries, we take the most recent value of the master report as the value of the project and derive changes to phases from the master and child reports. We consider a project to have started construction when the first child report has started construction. A project is completed when all the child projects complete, or when at least one child project has been completed and all other child reports have been abandoned (or have a pending deferral). We consider a project to have been abandoned when all child reports have been abandoned. The resulting data structure is such that every master-child grouping has only one entry for every date where data was available for the master report or one of the child reports.

We removed one project that matched with multiple master reports and its corresponding master reports from the data. There were sixteen projects for which a child report was made before the first parent report. In thirteen of these cases, the preceding child reports were all plans, in two cases a start report had been made for the child project before the first master project report was made, and in one case a completion report was made for the child project before the first master report entry. In all sixteen cases, we treat the history of the project as having started at the date of the first master project report entry and take the phase values from the child projects moving forward.

S.2.3. Project Removal

After a project has been reported on by Dodge, it may be removed from the set of projects that Dodge tracks for several reasons, including but not limited to:

1. The report is a duplicate report of another project.
2. The project's value fell below 500 thousand nominal US dollars.
3. The project was a master report and all of its child projects have received their first reports.
4. The project has not received a phase update in thirty-six months. Should the project later be updated, it will be re-introduced by Dodge to the data.

Dodge provides a removal date, which reflects the date at which Dodge will no longer provide updates to the project. Dodge also provides a comment on the reason for removal. We remove projects from the sample as of their removal dates, unless they are duplicate reports, in which case we remove all entries for that project from the data set.

Separately, we retain only project reports located in the United States.

Once the data cleaning described above is complete, we have a data set where we observe 3916 projects over time, with each project observed for 58 months on average, with no missing periods. All master and child report groupings are consolidated into a single series of project entries.