PitchBook

PitchBook Data, Inc.

Institutional Research Group

pbinstitutionalresearch@pitchbook.com

Published on November 14, 2025

Contents

PitchBook VC Valuation Estimates

PitchBook is a Morningstar company providing the most comprehensive, most accurate, and hard-to-find data for professionals doing business in the private markets.

Summary

PitchBook's VC-backed company valuation estimates—the expected pre-money valuation a company would be priced at under current market conditions—are derived from a regression model that predicts a company's expected annualized valuation growth rate from its last known post-money valuation.[1] The expected growth rates are based on three key sets of data sources: overall market growth, comparable company (comps) growth, and company-specific factors. We then calculate a nominal valuation estimate by applying the expected growth rate to a company's prior valuation.

$$\widehat{V}_t = V_{t-n}\left(1+\hat{g}_t\right)^{\frac{n}{365}}$$

Market and comp valuation-growth data is taken from two primary sources: public company equity performance and VC-backed company valuations observed in primary VC funding rounds. In addition to examining overall market performance from both sources, we generate a set of comps for each company being valued using a fine-tuned text embedding model introduced in this analyst note. The comp growth data for each source is then fed into the valuation model as an excess return-like figure relative to the respective market growth to help reduce redundancies across the inputs. Lastly, key company-specific inputs include employee growth, prior valuation growth, and age.

While predicting individual company growth rates is a challenging task, especially without access to detailed company data (such as financial statements), we found that the model inputs described above exhibited significant predictive power. When trained on our historical valuation growth data since 2005 and tested out-of-sample on data from 2021 through 2025, the model achieved an R^2 of 0.39, meaning that the inputs explained 39% of the variation across company valuation growth rates.[2] Predicting valuations in dollar terms is an easier task given that prior valuations provide a useful starting point. When applying the expected growth rates to calculate nominal valuations, the R^2 improved to 0.75. The median absolute percentage error was 30.8%, indicating that half of the predicted valuations were within ±30.8% of the actual valuation.

1: Valuation estimates are neither an indication of fair or intrinsic value nor what PitchBook believes a company is worth.
2: Model performance was validated via the step-ahead time series method in which predictions were made on data from a given calendar year from a model that was trained on data through the prior year-end.

Data

The historical valuation growth data used to train the model comes from PitchBook's database of global primary VC financing transactions. We focused on series-to-series step-up rounds—that is, Series A to Series B, Series B to Series C, and Series C to Series D—and growth rounds (Series D+) for the same company with a published valuation for both rounds, where growth was calculated based on the following equation.[3, 4]

$$g_t = \left(\frac{\text{Pre-money valuation}_t}{\text{Post-money valuation}_{t-n}} \right)^{\frac{365}{n}} - 1$$

This process yielded approximately 21,000 training observations of transactions completed between 2005 and October 2025. The following table offers additional details on the distribution of observations across stages.

Summary of annualized company valuation growth rate data used for training

Prior series	Series	Number of transactions	Average	Median	Lower quartile	Upper quartile
A	B	9,347	70.4%	51.1%	20.0%	105.2%
B	C	5,168	54.3%	37.4%	11.0%	81.8%
C	D	2,795	42.5%	27.7%	4.4%	68.5%
D+	D+	3,432	31.6%	17.3%	0.0%	49.8%

Source: PitchBook • Geography: Global • As of October 31, 2025
Note: Table includes data from financing rounds beginning in 2005.

The same valuation growth data that we trained the model to predict also serves as the source of the overall private market growth and private comp growth input data. The potential set of comparable transactions was determined based on the date differences of the ending valuation at time t and the starting valuation at time $t - n$. Lastly, public stock price data is sourced from Morningstar and includes formerly VC-backed companies that have traded on the New York Stock Exchange or Nasdaq.

Model inputs

Market growth

The private market growth factor (MGF) is the average growth observed across the full set of potential comps for a given stage (for example, a company that last raised a Series A would be compared with the observed growth of companies that raised a Series B).[5] The primary determinant of the potential comp set for company i at stage s is when it is being valued and when it was previously valued, denoted as t and $t - n$, respectively. Given that primary private valuations occur irregularly, it is not possible to obtain potential comp growth rates over the exact same period for each target

3: Only series-to-series rounds were considered so that we could group companies by maturity and fit different model parameters for each group. See the "Modeling" section for more details.
4: Growth rates calculated over periods of less than a year were not annualized. Only observations where n was less than 1,095 (three years) were used due to the increasing difficulty of predicting valuation growth over longer periods.
5: Annualized growth outliers of less than -70% and greater than 400% that could skew the average are excluded from the potential comp sets.

company. Therefore, we filtered the set of comps with respect to the overlap between the valuation growth period for the target company i and potential comp j. The valuation growth for company j observed between $t - n$ and t is included in the private MGF calculation for target company i if both of the following conditions are met:

The comparable transaction for which the valuation growth was calculated for company j occurred less than 180 days from when the target company is being valued: $t_i - t_j \leq 180$.

The prior valuation for the comparable transaction was dated within a year of the target company's last valuation: $|(t_i - n_i) - (t_j - n_j)| \leq 365$.

The potential comp set of private company growth rates (g_j) and the private MGF for company i with an expected next raise at stage s can then be formally defined as:

$$G_{i,\,\text{priv}} = \{g_j \in G_{\text{priv}} \mid s_i = s_j, t_i - t_j \leq 180, |(t_i - n_i) - (t_j - n_j)| \leq 365\}$$

$$\text{MGF}_{i,\text{priv}} = \frac{1}{|G_{i,\text{priv}}|} \sum_{g_j \in G_{i,\text{priv}}} g_j$$

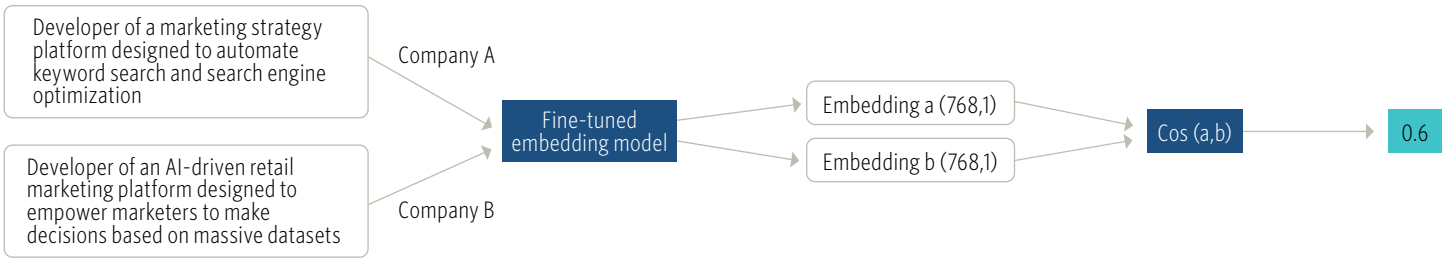
The public MGF is simply the change in the Morningstar US Broad Growth Extended Index over the period of $t - n$ to t for company i that is being valued at time t and was valued previously at time $t - n$.

$$\text{MGF}_{i,\text{pub}} = \left(\frac{\text{Index}_{t_i}}{\text{Index}_{t_i - n_i}} \right)^{\frac{n_i}{365}} - 1$$

Comparable company growth

Comp growth factors (CGFs) represent the average valuation or stock price growth of a subset of the potential comp sets based on company similarity for each of the two growth data sources. CGFs are calculated as an excess return metric relative to the MGFs for each respective price source, thereby removing redundant market performance information. This step can be thought of as separating market and comp-specific effects, which can then be distinctly parameterized during modeling. The measure of company similarity between target company i and each potential comp j is automated at scale via a fine-tuned transformer embedding model, which is fed company descriptions and keywords for pairs of companies and outputs a similarity score from -1 to 1.

Illustrative example of the company similarity modeling process



Source: PitchBook

Fine-tuning a pretrained semantic similarity embedding model is a challenging yet essential step, as it involves measuring the similarity between pairs of text, which is inherently subjective. We used our analyst-curated technology verticals (ACVs) developed by our Industry and Technology Research team to establish baseline expectations of similarity between pairs of startups. These similarity expectations were then employed to generate labeled data pairs ranging from 0 to 1 for fine-tuning the pretrained similarity embedding model.

At the time this model was fine-tuned, there were 26 ACVs, each sitting at the top of a three-tiered hierarchical taxonomy. Verticals contain segments, and each segment is further divided into subsegments. This three-tiered taxonomy allowed us to create four groups of company pairs for labeling purposes: companies that are not in the same vertical, companies that are in the same vertical but different segments, companies that are in the same segment but different subsegments, and companies that are in the same subsegment. We then assigned expected similarity scores based on which of these four groups a pair of companies fell into and the simple assumption that similarity should monotonically increase as the groups narrow. Companies in different verticals are expected to be the least similar, and companies within the same subsegment are expected to be the most similar. Lastly, we assumed that the expected similarity scores increase linearly from one group to the next over the range from 0.2 to 0.8.

For target company i and each potential comp j in G_i, the model outputs a similarity score, $s_{i,j}$. This score is then used to filter G_i into the final comp set before calculating the CGF, which we denote as C_i. Based on the fine-tuning process, we set the threshold for the similarity score at 0.4, which is roughly equivalent to companies in the comp set being in the same vertical as the target company. Further, to increase the concentration of the comp set and differentiate the CGF from its respective MGF, we consider only the top 50 comps ranked by similarity score if C_i contains more than 50 comps. Additionally, we required at least 10 comps to calculate the CGF. The following procedure applies to both growth data sources.

$$C_i = \{g_j \in G_i | s_{i,j} > 0.4\}$$

$$\mathrm{CGF_i} = \frac{1}{|\mathrm{Top}_{50}(C_i)|} \sum_{g_j \in \mathrm{Top}_{50}(C_i)} g_j - \mathrm{MGF}_i, \ if\ |C_i| \geq 10$$

Company

Employee growth is the key company-specific input to the model, which we use as a proxy for individual company performance. While employee growth is somewhat limited in this capacity, it is much more widely available than financial data for privately held companies. Because PitchBook captures employee figures at irregular intervals, we created a rules-based process to determine which data, if any, to include in the employee growth calculation for company i that is being valued at time $t_{i,\text{val}}$. The following rules are applied to determine the employee growth input, $g_{i,\,\text{emp}}$:

> The more recent employee count figure, $c_{i,t}$, must be from within six months of $t_{i,\text{val}}$ with a preference for the most recent one if multiple data points qualify: $0 \leq t_{i,\,\text{val}} - t \leq 180$.
>
> The prior employee count, $c_{i,t-n}$, must be from at least six months before the more recent employee count and no more than a year before with a preference for the most recent one if multiple data points qualify: $180 \leq n \leq 365$.

This process results in employee growth inputs over a trailing six- to 12-month window that are from within six months of the respective valuation estimate date. Company-specific employee growth rates are calculated relative to the average employee growth observed at each stage, which is also included as a separate input. In addition to employee growth, other company-specific inputs include company age (time since founding), time since a company's last VC financing round, and the company's prior growth rate.

Summary of model inputs

	Input	Description
Market	Public market growth factor	The price return of the Morningstar US Broad Growth Extended Index from when a company was previously valued to the valuation estimate date.
	Private market growth factor	The average private company valuation growth rate derived from relevant primary transactions based on stage and time.
	Average employee growth	The average employee growth rate of relevant VC-backed companies based on stage and time.
Comps	Public comp growth factor	The average excess price return of relevant, formerly VC-backed companies that are publicly traded based on similarity scores.
	Private comp growth factor	The average excess valuation growth of relevant VC-backed companies based on stage and similarity scores.
Company-specific	Previous annualized valuation change	The last known growth rate of a company observed at a primary VC funding round.
	Relative employee growth	A company's employee growth rate relative to average employee growth.
	Company age	Number of years since founding (or first VC deal if the founding date is unknown).
	Days since last value	Number of days since the company was previously valued during a primary VC funding round.

Source: PitchBook

Note: For estimating current valuations, days since last value is imputed, given that the training labels are observed only when a transaction occurs.

Modeling

We modeled annualized growth rates as a lognormal variable, $(1 + g) \sim \text{LogNormal}(\mu_g, \sigma_g)$, where μ_g and σ_g were both determined via linear regression.[6] We found that this simple model structure performed nearly as well as more complex models, such as a neural network, and provided much better interpretability of the model and its predictions. The model was fit via Bayesian inference, which allows us to produce predictions that are full probability distributions and provide a clearer understanding of uncertainty.[7] Unlike a traditional regression that fits parameters by minimizing the mean squared error, a Bayesian regression requires declaring prior probability distribution assumptions (that is, priors) for each parameter, which are then updated given observed data. These updated distributions are referred to as the posterior distribution, formally defined as $\text{P}(\theta \mid \text{Data})$.

Another benefit of Bayesian modeling is that it naturally handles hierarchical models within a single structure rather than requiring separate models with independent parameters for each group. In this case, we structured a hierarchical model into four groups based on company stage transitions: Series A to Series B, Series B to Series C, Series C to Series D, and Series D+ (Series D to Series E, Series E to Series F, and so on). While all four groups have their own regression parameters, their prior distributions share a common global mean, which effectively allows information sharing across groups. The full model structure is defined below, where X and Z represent the mean-variance standardized input data for the mean and sigma regressions, respectively. The sigma regression inputs include the cross-sectional standard deviation of private market growth, company age, time since last raise, and a binary indicator of whether each observation has a known employee growth figure.

Global priors

$$\alpha_{\text{global}}, \beta_{\text{global}}, \gamma_{\text{global}} \sim \text{Normal}(0, 1)$$

Mean regression priors

$$\alpha_s \sim \text{Normal}(\alpha_{\text{global}},\ 1)$$

$$\beta_s \sim \text{Normal}(\beta_{\text{global}}, 0.2)$$

Sigma regression priors

$$\gamma_{s,0} \sim \text{Normal}(0, 1)$$

$$\gamma_s \sim \text{Normal}(\gamma_{\text{global}}, 1)$$

Regressions

$$\mu_g = \alpha_s + \beta_s \text{X}$$

$$\sigma_g = \exp(\gamma_{s,0} + \gamma_s Z)$$

6: To mitigate the high degree of skew in the transformed raw growth distribution that is produced from the lognormal assumption, we truncated the lognormal distribution at log(4), which leads to better calibrated predictions.
7: The model was implemented via the Python package PyMC using the No-U-Turn Sampler (NUTS) Hamiltonian Monte Carlo (HMC) method.

Likelihood

$$(1+g) \sim \text{LogNormal}(\mu_g, \sigma_g)$$

Derived valuation distribution

$$V_t \sim \text{LogNormal}\left(\log V_{t-n} + \mu_g \frac{n}{365},\ \sigma_g \frac{n}{365}\right)$$

A summary of the model parameters' posterior distributions broken out by stage is shown below. In general, we found the resulting parameters' implied relationship with valuation growth to be intuitive. For example, both higher MGFs and CGFs lead to a higher estimated growth rate; higher employee growth results in a higher estimated growth rate; older companies and those that have taken longer to raise have lower expected growth rates; and companies with strong prior valuation growth have higher expected growth rates. Additionally, with respect to the sigma model, we found that greater cross-sectional variance in private market growth rates and time since last value increased expected uncertainty, while increases in company age and having an employee growth figure decreased expected uncertainty.

Average log growth model coefficients by stage

	Series A to B	Series B to C	Series C to D+	Series D+ to D+
Intercept	0.40 [0.39, 0.41]	0.37 [0.36, 0.38]	0.35 [0.33, 0.37]	0.34 [0.32, 0.37]
Public MGF	0.03 [0.02, 0.04]	0.04 [0.03, 0.05]	0.05 [0.04, 0.07]	0.04 [0.02, 0.05]
Private MGF	0.08 [0.07, 0.10]	0.09 [0.08, 0.11]	0.09 [0.07, 0.11]	0.11 [0.09, 0.12]
Average employee growth	0.03 [0.02, 0.04]	0.02 [0.01, 0.03]	0.02 [0.00, 0.04]	0.01 [0.00, 0.03]
Public CGF	0.00 [0.00, 0.01]	0.01 [0.00, 0.02]	0.02 [0.00, 0.03]	0.01 [0.00, 0.02]
Private CGF	0.05 [0.04, 0.06]	0.05 [0.04, 0.06]	0.04 [0.02, 0.06]	0.05 [0.03, 0.06]
Previous growth	0.04 [0.02, 0.06]	0.06 [0.05, 0.07]	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]
Relative employee growth	0.07 [0.06, 0.08]	0.08 [0.06, 0.10]	0.08 [0.05, 0.10]	0.08 [0.05, 0.10]
Company age	-0.06 [-0.07, -0.05]	-0.03 [-0.05, -0.02]	-0.03 [-0.04, -0.02]	-0.03 [-0.04, -0.02]
Time since last value	-0.15 [-0.16, -0.15]	-0.14 [-0.15, -0.13]	-0.07 [-0.08, -0.06]	-0.07 [-0.08, -0.06]

Source: PitchBook

Note: Coefficients priors are standard normal and are applied to mean and variance standardized input values. The bracketed range contains the 95% high-density interval.

Evaluation

The model was evaluated through annual step-ahead time series validation, starting with the first validation set in 2021 and concluding with the last one in 2025. Point estimation performance metrics were calculated based on the medians of the posterior distributions, and the reported statistics were averaged across the step-ahead validation sets. With respect to annualized growth rates, the model achieved an out-of-sample R^2 of 0.39 and a median absolute error of 28.1% across all stages.[8] When the estimated growth rates were applied to the last known valuations (V_{t-n}) to get the current valuation estimates $(\widehat{V_t})$, the R^2 improved to 0.75 and the median absolute percentage error (MAPE) was 30.8%. Absolute model performance improved as companies aged because growth rates tend to be less volatile for more mature companies, with the MAPE ranging from 31.5% to 24.6% from Series A to Series D+, respectively. Given the relatively small sample sizes at the stage level in the validation sets, these figures were calculated from the full sample.

Out-of-sample model performance summary by stage

			Valuation growth		Nominal valuation	
Prior series	Series	Total number of observations	Median absolute error	R^2	Median absolute percentage error	R^2
A	B	2,695	30.5%	0.38	30.6%	0.71
B	C	1,420	30.1%	0.39	33.4%	0.73
C	D	790	28.9%	0.39	32.2%	0.71
D+	D+	1,080	21.9%	0.36	26.8%	0.73
Overall		5,990	28.1%	0.39	30.8%	0.75

Source: PitchBook

Note: Performance metrics are averaged across the annual step-ahead validation sets from 2021 to 2025.

In-sample model performance summary by stage

			Valuation growth		Nominal valuation	
Prior series	Series	Total number of observations	Median absolute error	R^2	Median absolute percentage error	R^2
A	B	9,347	33.2%	0.38	31.5%	0.74
B	C	5,168	28.7%	0.40	30.9%	0.78
C	D	2,795	26.5%	0.40	26.5%	0.76
D+	D+	3,432	21.3%	0.37	24.6%	0.76
Overall		20,742	28.8%	0.40	29.8%	0.77

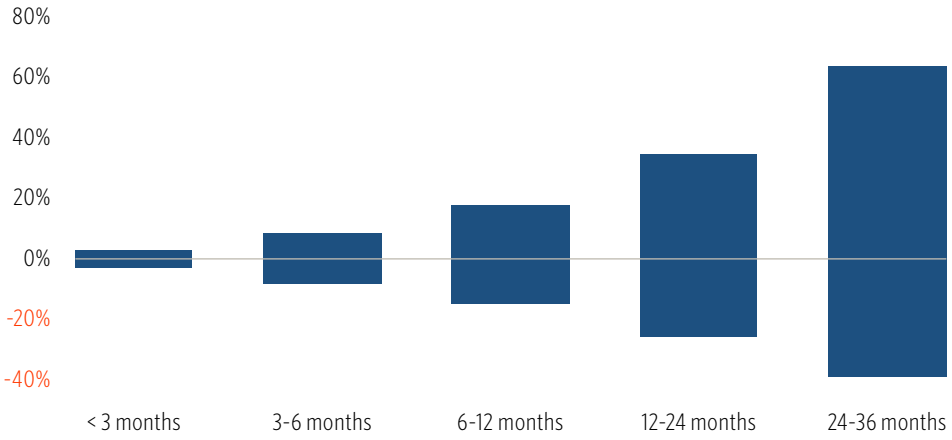
Source: PitchBook

8: The reported R^2 figures were calculated via a method designed for Bayesian models defined in "R-Squared for Bayesian Regression Models," Columbia University, Andrew Gelman, et al., November 4, 2018.

While the model has statistically significant predictive power, it is important to note that a considerable amount of unexplained variance remains in the valuation estimates. This phenomenon is primarily driven by the fact that VC-backed company growth rates are highly volatile, even when conditioned on the same set of market conditions.

Time is another key consideration when assessing model performance as well as individual company valuation estimates. Because the growth rate is exponentiated in the valuation estimation equation $\left(\widehat{V}_t = V_{t-n}(1+\hat{g}_t)^{\frac{n}{365}}\right)$ and $1+\hat{g}_t$ is a lognormal random variable, $\widehat{V}_t$ is also a lognormal random variable. Further, the variance of $\widehat{V}_t$ will be equal to the variance of the growth rate scaled by the square of time in annual units. For example, consider two companies with the same expected annualized growth rate variance, but the first was last valued one year ago and the second was last valued three years ago. The variance of the nominal valuation estimate for the second company will be $9\text{x}\left(3^2/1^2\right)$ that of the variance of the valuation estimate for the first company. Intuitively, then, the model produces more confident predictions over shorter time horizons.

Interquartile range of current valuation estimate distributions by time since last valuation



Source: PitchBook

Note: Valuation estimates were generated on October 31, 2025. The values represent the percentage difference from the median.

Estimations

Eligibility

There are three requirements for a company to receive a valuation estimate on the PitchBook Platform:

- The company must be currently VC backed with a closed Series A or later round.[9]
- The company must have a post-money valuation on one of its previous VC rounds (Series A or later).
- The company's latest post-money valuation must be within the past three years.

Feature attribution

There are several challenges to additively explaining individual valuation estimates in terms of the model inputs, despite the model being linear. First, although the log growth model underpinning the valuation estimates is linear, there are two nonlinear transformations that take place to derive the nominal valuation estimate: 1) exponentiating log growth and 2) exponentiating $(1+\widehat{g})$ by time to get the unannualized valuation factor by which to multiply the previous valuation. Applying these same transformations to $c_j = x_j\beta_j$ for each feature j will not yield additive feature contributions. Second, it is not possible to directly calculate the expected median valuation of the posterior distributions from the posterior distributions of the log growth regression parameters.

To appropriately handle the nonlinear aspects of the model, we employ Kernel SHAP,[10] a model-agnostic attribution method that approximates Shapley values by feature coalition sampling and local weighted regressions. We define c_j as each feature's marginal contribution to the nominal difference between the median posterior valuation and the prior valuation: $\widehat{V}_t - V_{t-1}$. For a given feature coalition, S, posterior samples of the growth regression parameters α and β, and features values x, we calculate the following values.

$$\mu_g = \alpha + \beta \mathrm{x}^T$$

$$\widehat{\mathrm{V}}_t = V_{t-1} \cdot \exp\left(\mu_g \cdot \frac{n}{365}\right)$$

$$\Delta\widehat{\mathrm{V}}_t = \widehat{\mathrm{V}}_t - V_{t-1}$$

$$\Delta\widehat{V}_t = \mathrm{median}\left(\Delta\widehat{\mathrm{V}}_t\right)$$

9 : If a company has a generically tagged VC deal type of "early-stage VC" or "late-stage VC," we will impute its associated stock type based on the company's prior deal history and the simple assumption that companies progress through stages linearly.

10: "A Unified Approach to Interpreting Model Predictions," arXiv, Scott M. Lundberg and Su-In Lee, November 25, 2017.

The Kernel SHAP algorithm then estimates each feature's marginal contribution to the predicted change in nominal valuation, c_j, as shown below.

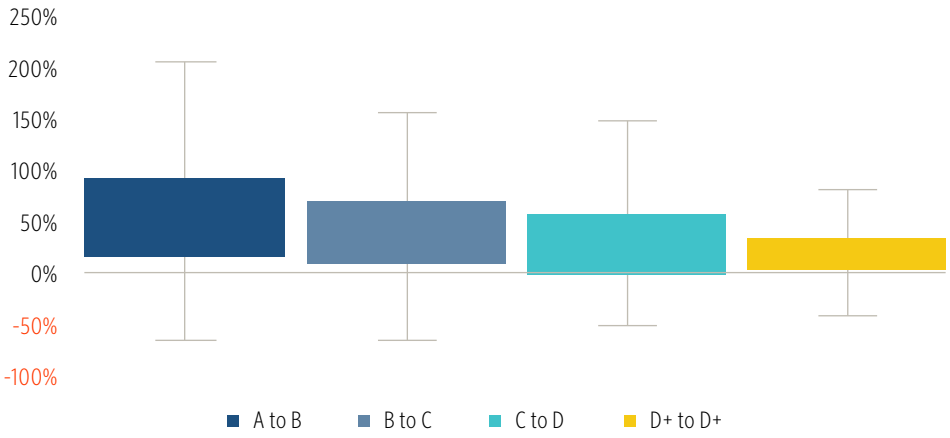
$$E\left[\widehat{V}_t \middle| \mathbf{x} = 0\right] + \sum_{j=1}^{J} c_j = \Delta\widehat{V}_t$$

Lastly, we calculate the gap between the median valuation estimate derived from the log growth regression parameters and the median valuation estimate from the actual posterior distribution. Given that the posterior distribution is a mixture distribution of D draws, these two figures will not be equal. However, because they rarely exceed ±2% of the prior valuation, we incorporate this gap into the expected value term to ensure equivalence between the left- and right-hand sides of the equation above.

Limitations and assumptions

While the model provides a structured and objective approach to estimating VC-backed company valuations, it is important to acknowledge several limitations and assumptions that influence how the estimates are used and interpreted. First and foremost, there is a naturally high degree of variation in startup growth rates, which are significantly influenced by idiosyncratic factors. The following chart visually illustrates this concept by displaying the distributions of growth rates across stages for companies that have raised capital since 2024 under similar market conditions. For companies that raised a Series C, for example, the interquartile range of the distribution alone spans from 5.9% to 65.9%, and the middle 90% range is from -68.1% to 153.2%.

Dispersion of annualized growth rates for companies that have raised capital since 2024



Source: PitchBook • Geography: Global • As of October 31, 2025

A second key limitation of the model is that it incorporates incomplete company-specific signals due to the lack of detailed financial disclosures and standardized reporting for private companies. While employee growth serves as a proxy for underlying business momentum, it is unlikely to capture the full picture for most companies. Therefore, we view these valuation estimates as primarily a mark-to-market exercise that is complemented with some idiosyncratic adjustments.

In addition to these key limitations, several assumptions in the modeling process are worth highlighting. The training data is constructed from historical valuations recorded at the time companies raise new funding rounds, which embed several assumptions in the predictions. First, we inherently assume that the companies being valued will raise more capital either privately or publicly (that is, the companies are going concerns). This will cause the valuation estimates to be overly optimistic in aggregate because they ignore out-of-business risks. Another embedded assumption is that growth between rounds is linear, given that we do not observe how valuations progress in the interim. This introduces an unknown source of error when valuing companies in between rounds because startup growth is likely nonlinear and path dependent.