Simulation-Informed Revenue Extrapolation with Confidence Estimate for Scaleup Companies Using Scarce Time-Series Data

Lele Cao

Sonja Horn

Vilhelm von Ehrenheim

Richard Anselmo Stahl

Henrik Landgren

MOTHERBRAIN

Revenue and Scaleup

Revenue: total income from generated from main business, indicating performance of a company's performance.

Scaleups: companies with proven scalability, viability and accelerated revenue growth.

Revenue is a highly relevant metric to evaluate a scaleup company!

Revenue Forecast

Investment professionals(IP) rely on **extrapolating company revenue** into the future to **approximate the valuation of companies** and inform their investment decision

Financial data on scaleups is typically **proprietary, costly and scarce,** forming a huge **obstacle** for directly **applying data-driven methodologies**

Forecasting typically done **manually** and **empirically** leaving the quality **heavily dependent** on the investment professionals' experiences and insights

Promise of Data-Driven Approach

Level of **automation, objectiveness, consistency** and **adaptability** for empirical revenue forecasting is **far from optimal**

Highly desirable for investment professionals evaluating scaleups to have a data-driven method that performs revenue extrapolation on scarce data in an automated way

- A **quick way** to **assess companies' revenue potential** with little information needed
- **Benchmarking** of a **manually produced revenue forecasting**

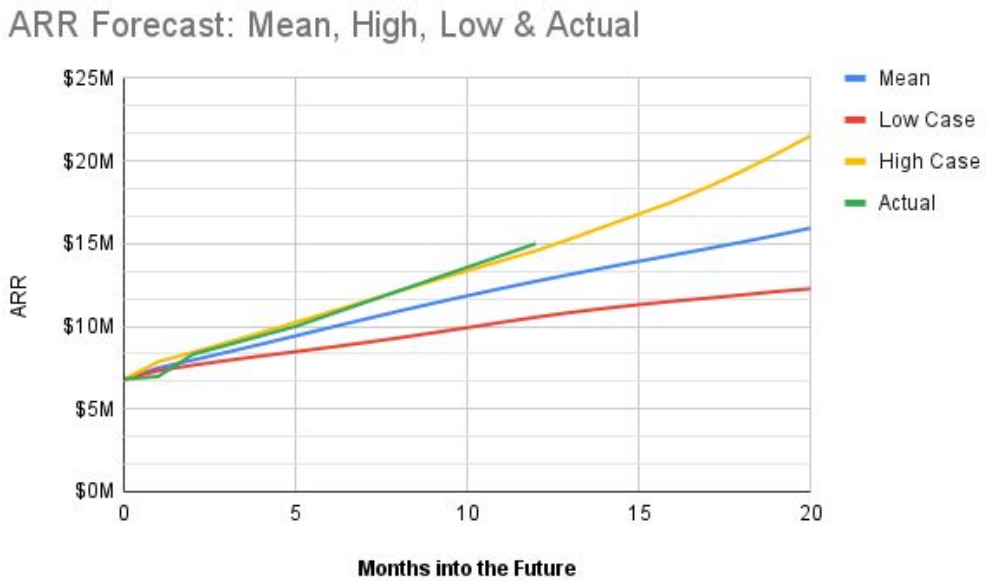
Data-Driven Revenue Forecast

The algorithm should

- work for multiple business sectors,
- work on a small dataset,
- commence from short time-series,
- extrapolate for long term (e.g. 3 years),
- estimate confidence,
- have low requirement on auxiliary information,
- be easy to explain.

This is the first work that meets all practical requirements simultaneously.

Example


ARR Forecast: Mean, High, Low & Actual
$25M
$20M
$15M
$10M
$5M
$0M
ARR
0
5
10
15
20
Months into the Future
Mean
Low Case
High Case
Actual

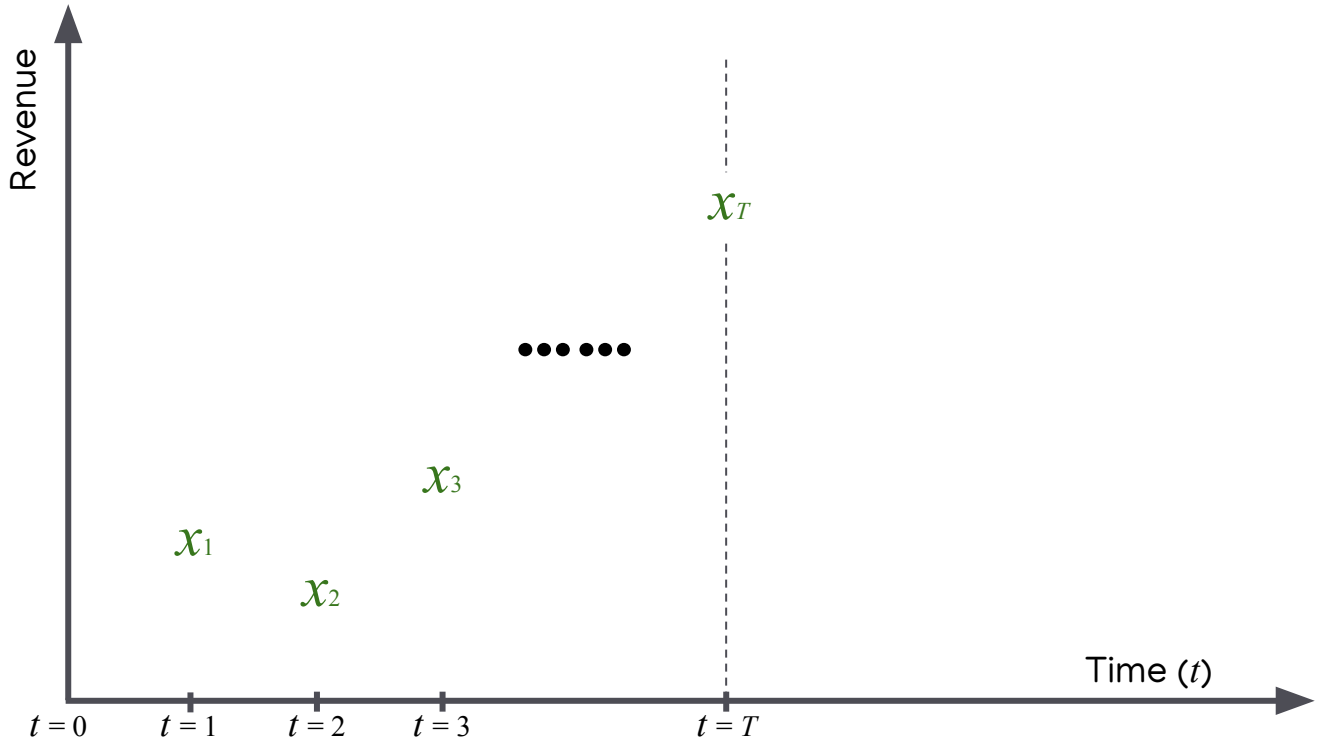
Revenue Model: Notation

x_t The true unobserved revenue

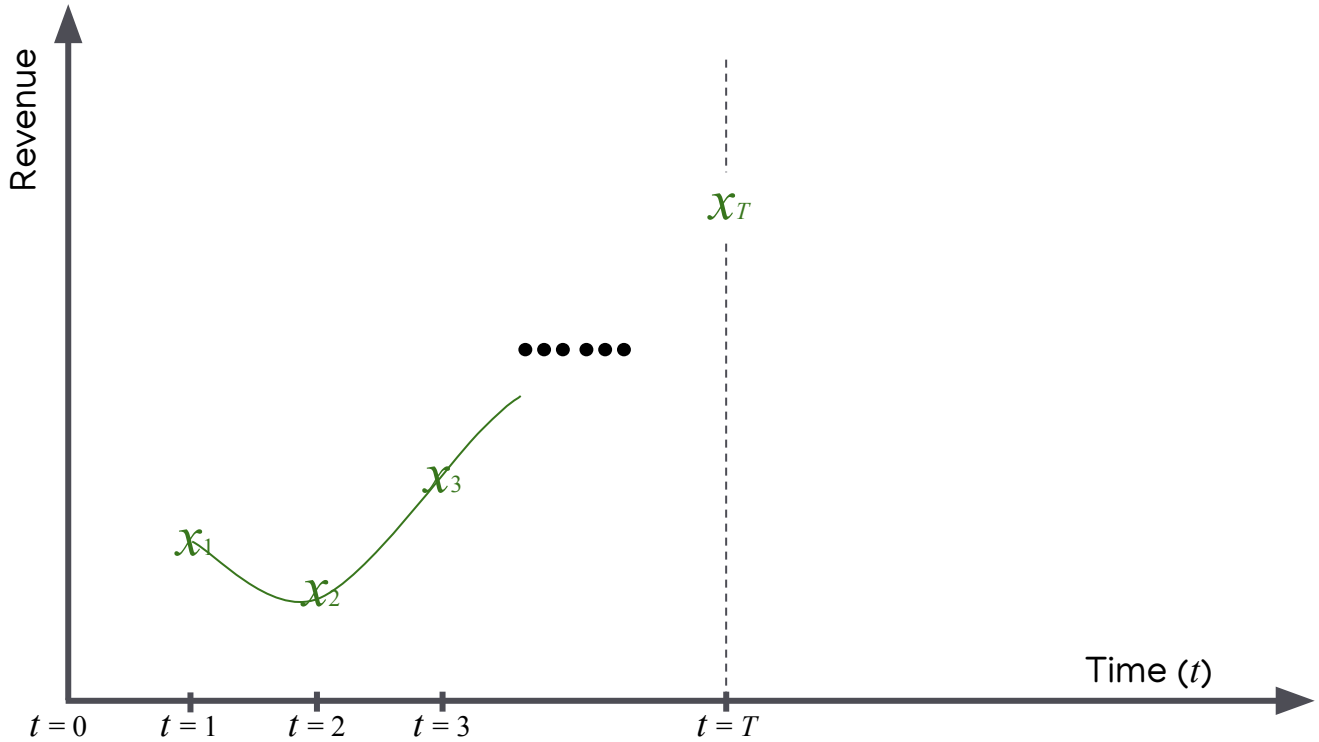
y_t The “noisy” measurements obtained through estimation

u_t The observed “booked” historical revenue numbers

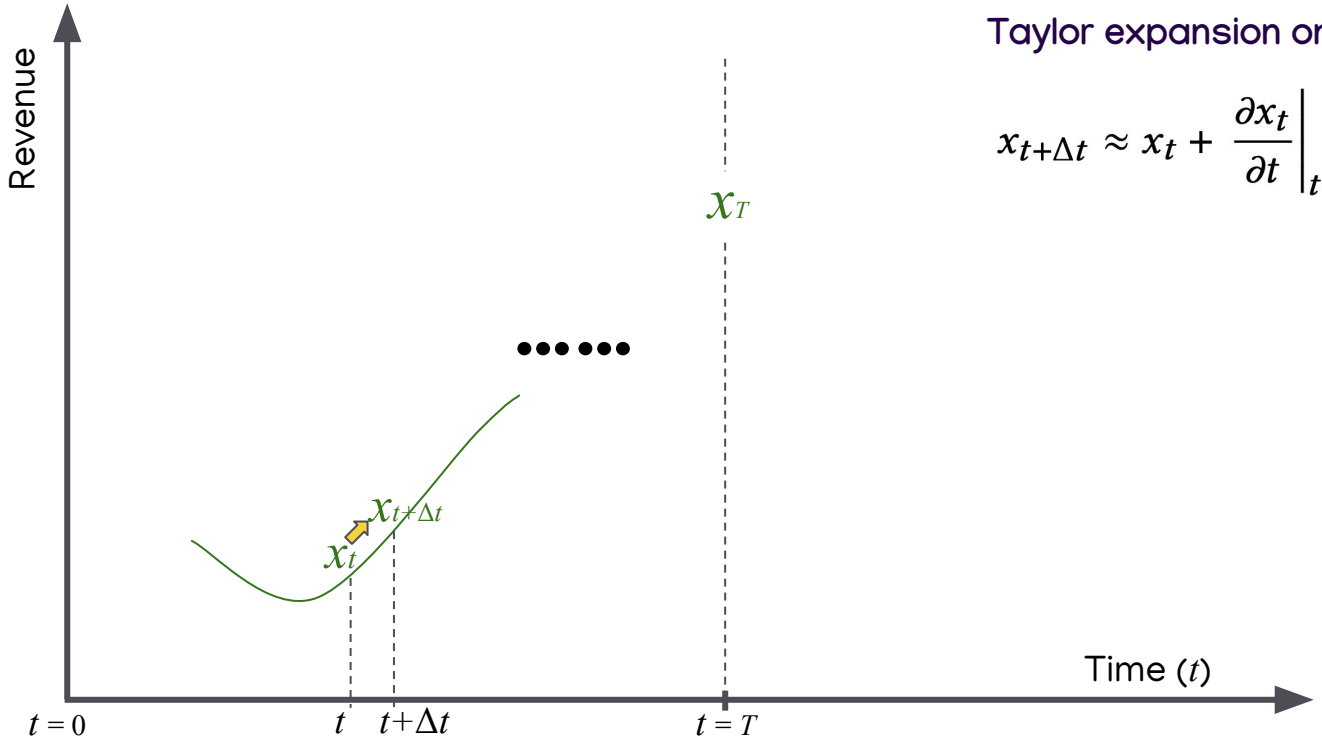
Revenue Model



Revenue Model



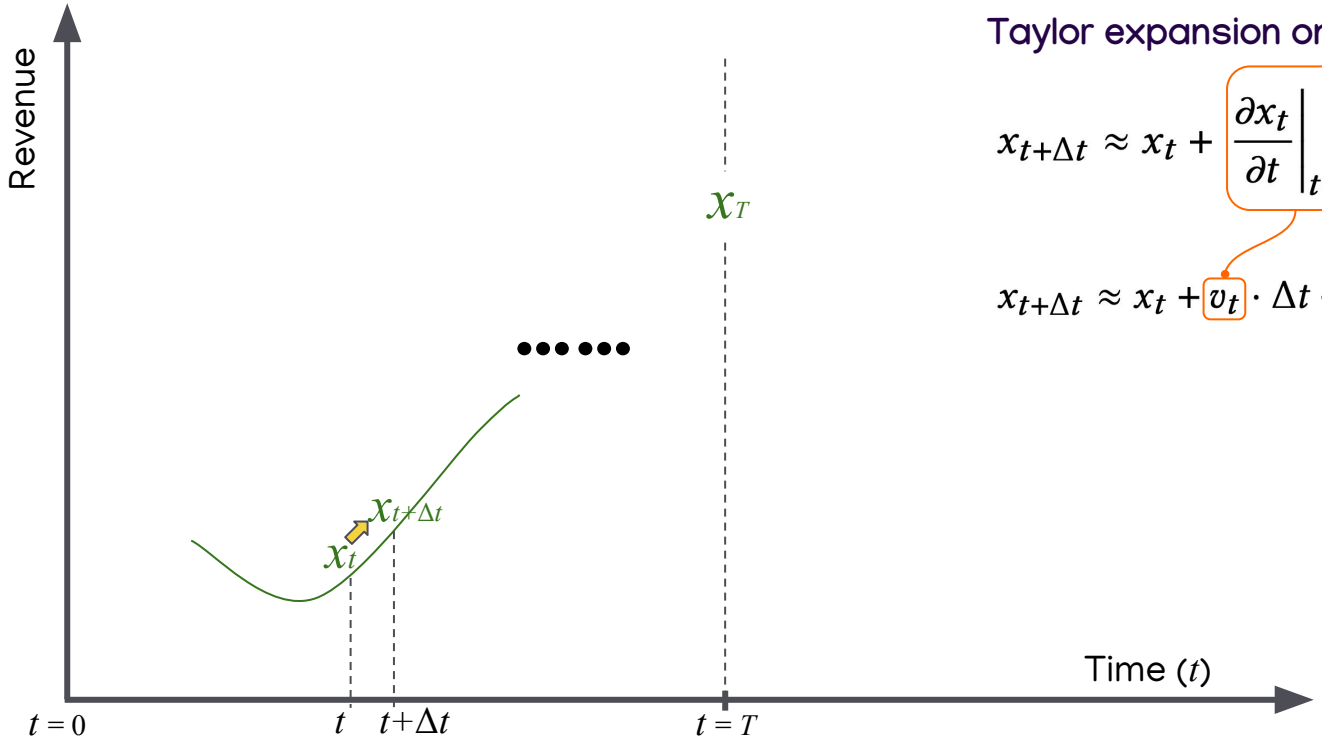
Revenue Model



Taylor expansion on x :

$$x_{t+\Delta t} \approx x_t + \left.\frac{\partial x_t}{\partial t}\right|_{t=0} \cdot \Delta t + \frac{1}{2} \left.\frac{\partial^2 x_t}{\partial t^2}\right|_{t=0} \cdot \Delta t^2$$

Revenue Model

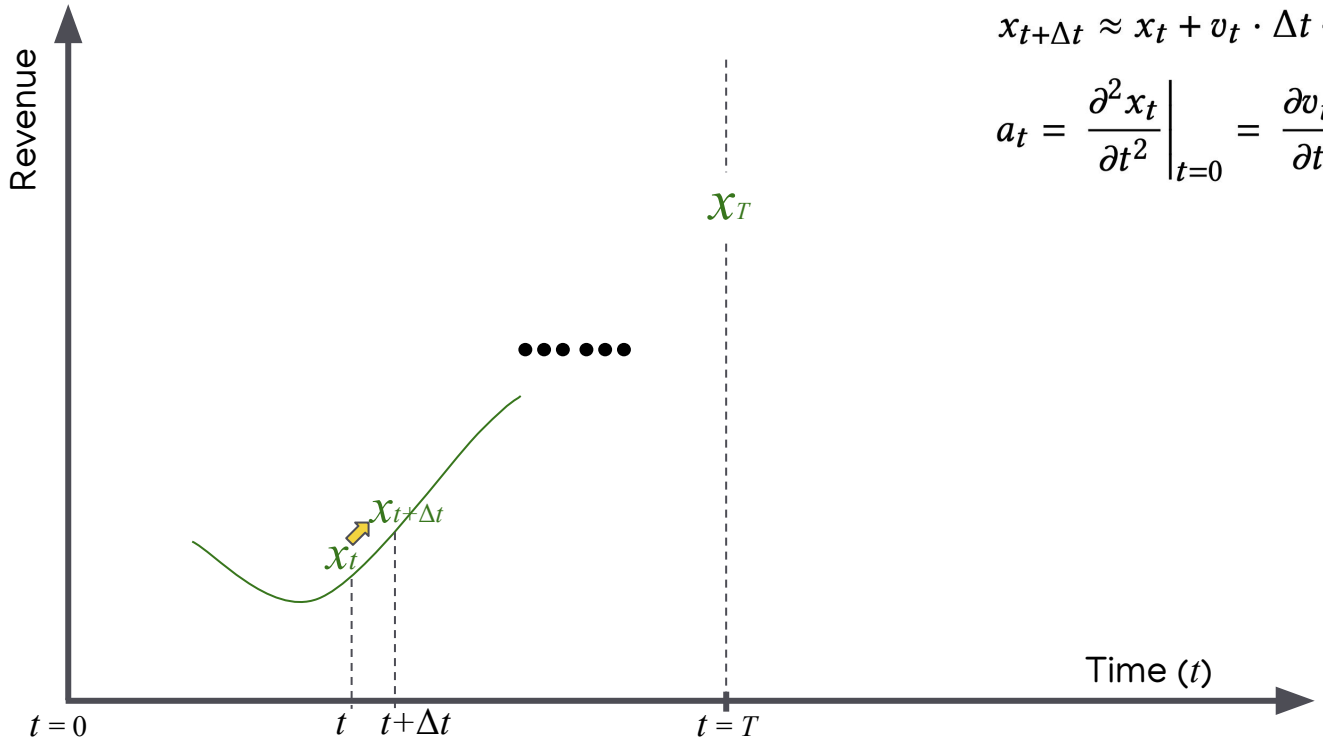


Taylor expansion on x :

$$x_{t+\Delta t} \approx x_t + \left.\frac{\partial x_t}{\partial t}\right|_{t=0} \cdot \Delta t + \frac{1}{2}\left.\frac{\partial^2 x_t}{\partial t^2}\right|_{t=0} \cdot \Delta t^2$$

$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

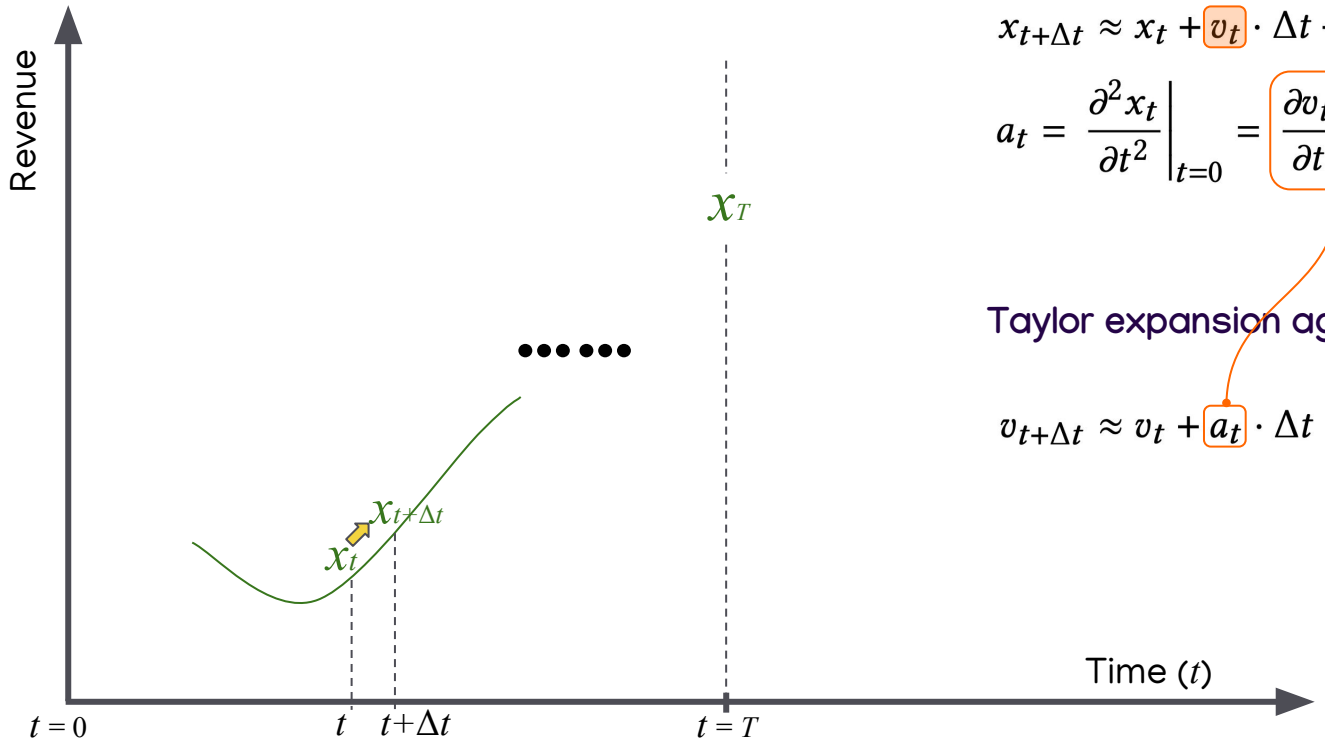
Revenue Model



$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$a_t = \left.\frac{\partial^2 x_t}{\partial t^2}\right|_{t=0} = \left.\frac{\partial v_t}{\partial t}\right|_{t=0}$$

Revenue Model



$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$a_t = \left.\frac{\partial^2 x_t}{\partial t^2}\right|_{t=0} = \left.\frac{\partial v_t}{\partial t}\right|_{t=0}$$

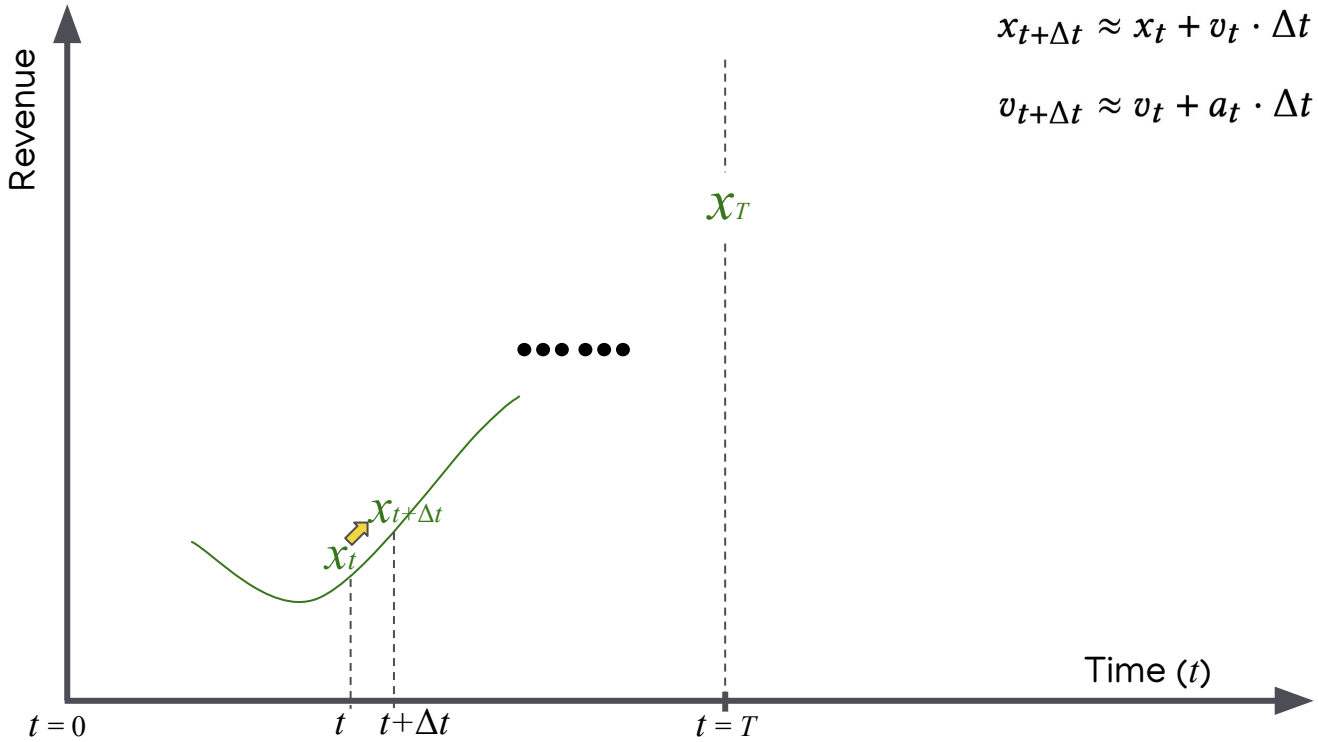
Taylor expansion again on v :

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

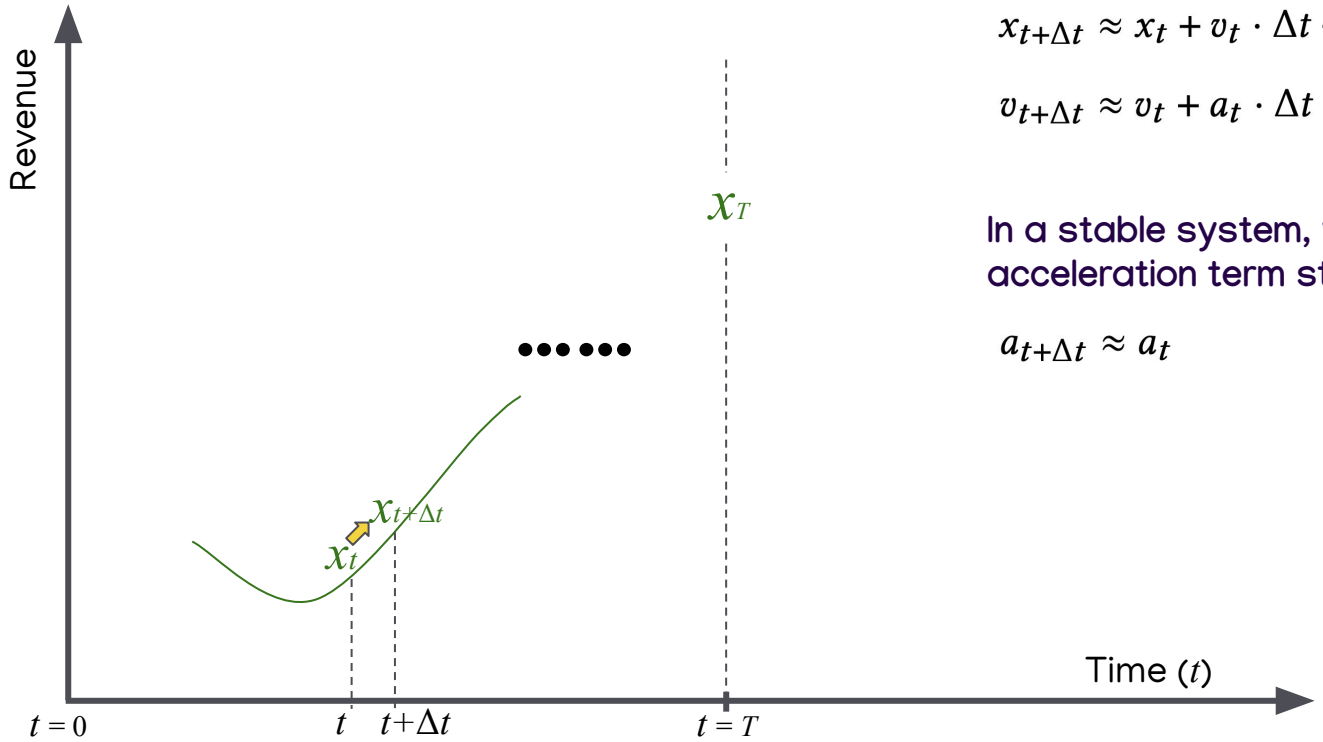
Revenue Model

$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$



Revenue Model



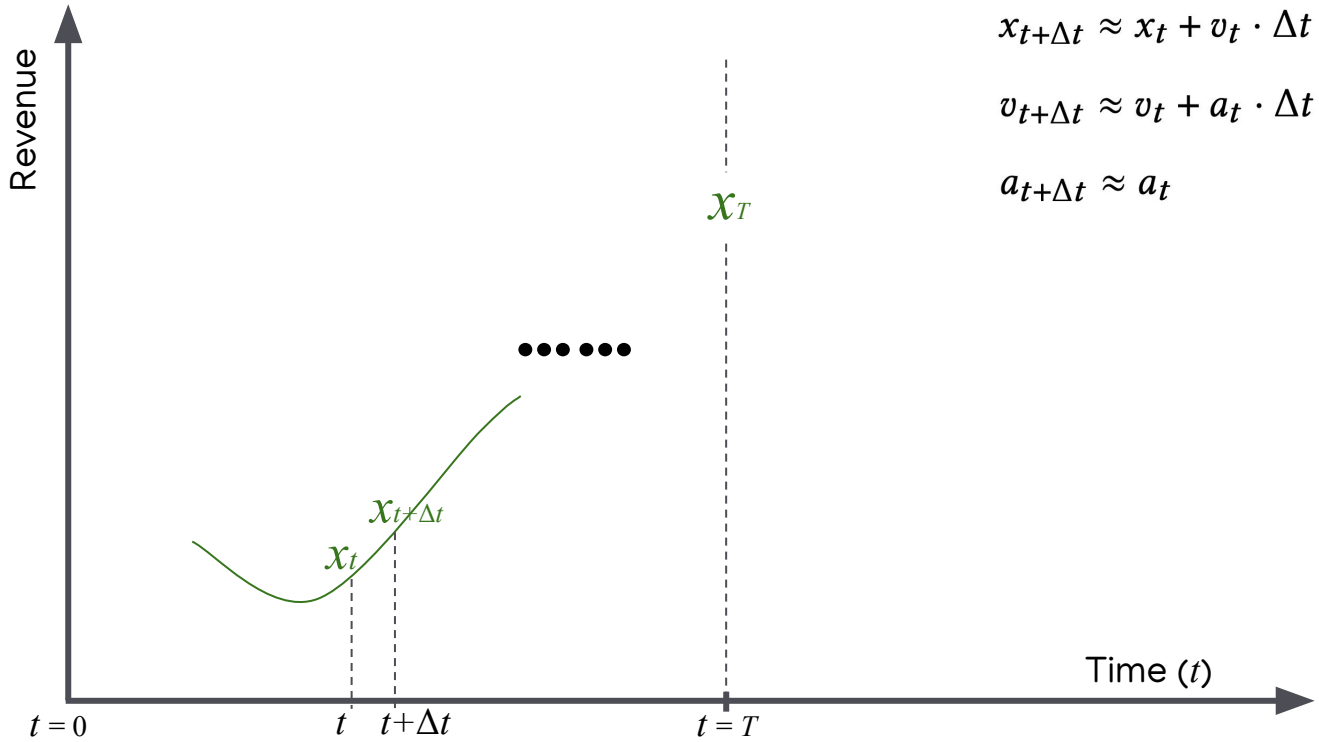
$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

In a stable system, we may assume that the acceleration term stays largely constant :

$$a_{t+\Delta t} \approx a_t$$

Revenue Model

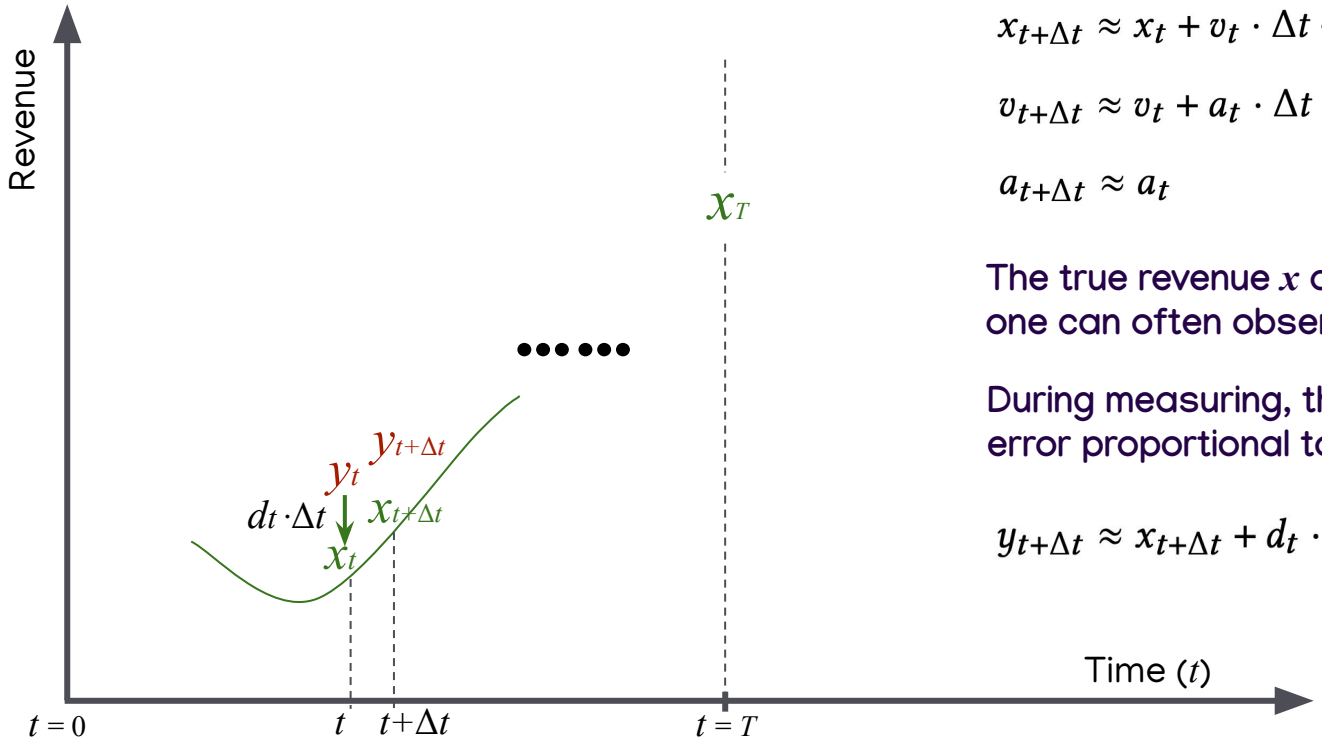


$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

$$a_{t+\Delta t} \approx a_t$$

Revenue Model



$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

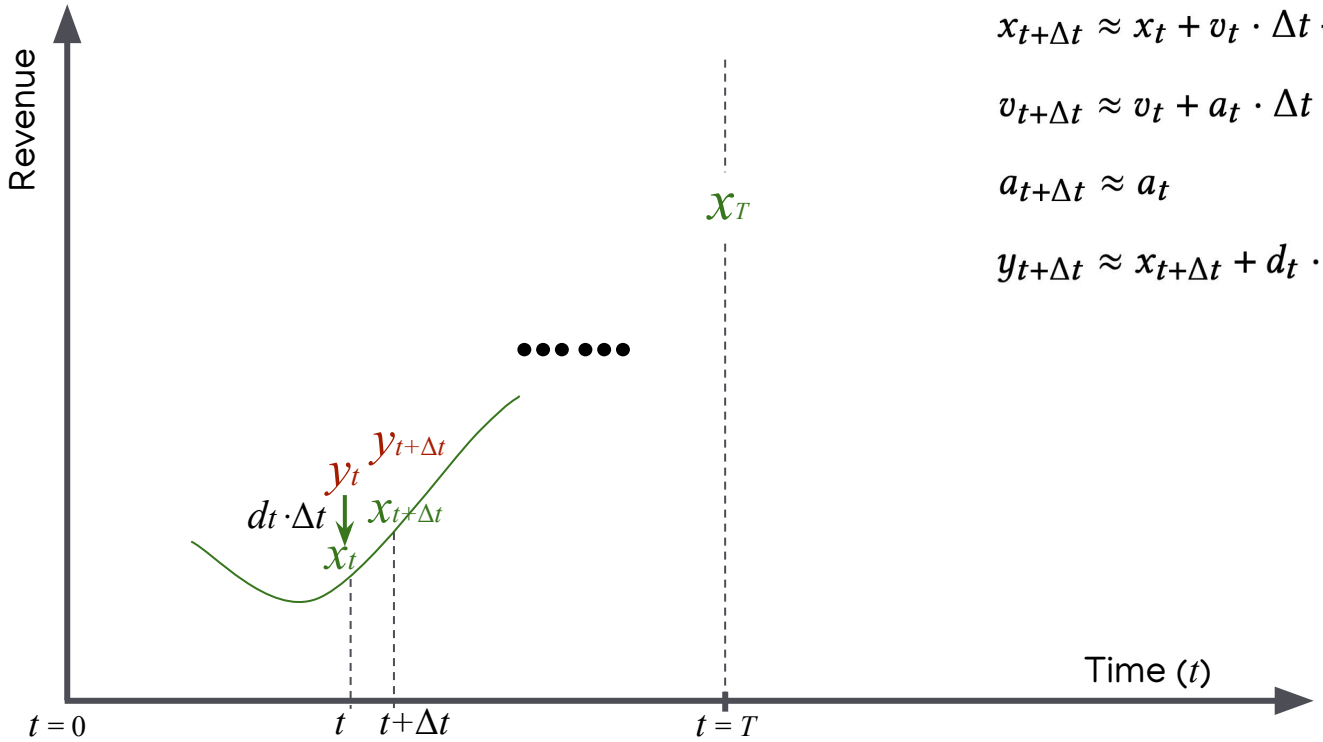
$$a_{t+\Delta t} \approx a_t$$

The true revenue x are usually "hidden", yet one can often observe the measured value y.

During measuring, there can be a systematic error proportional to Δt

$$y_{t+\Delta t} \approx x_{t+\Delta t} + d_t \cdot \Delta t$$

Revenue Model



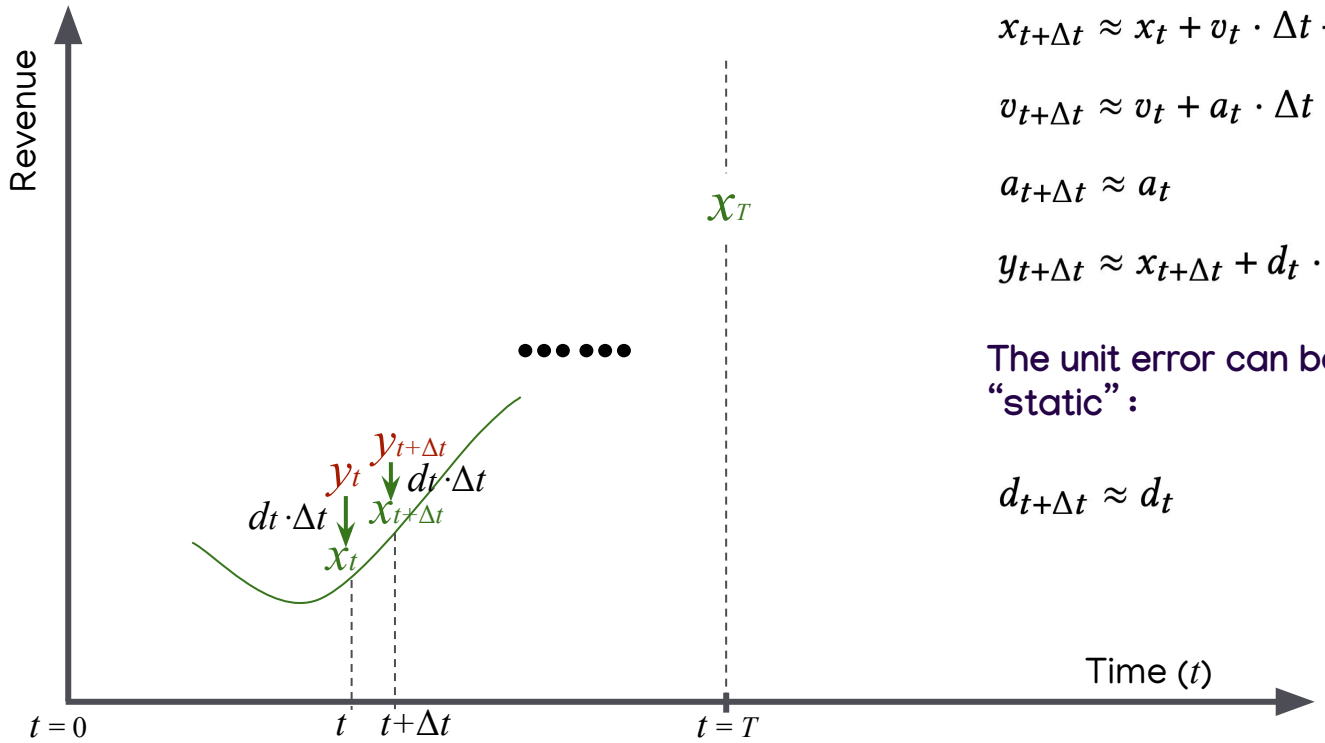
$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

$$a_{t+\Delta t} \approx a_t$$

$$y_{t+\Delta t} \approx x_{t+\Delta t} + d_t \cdot \Delta t$$

Revenue Model



$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

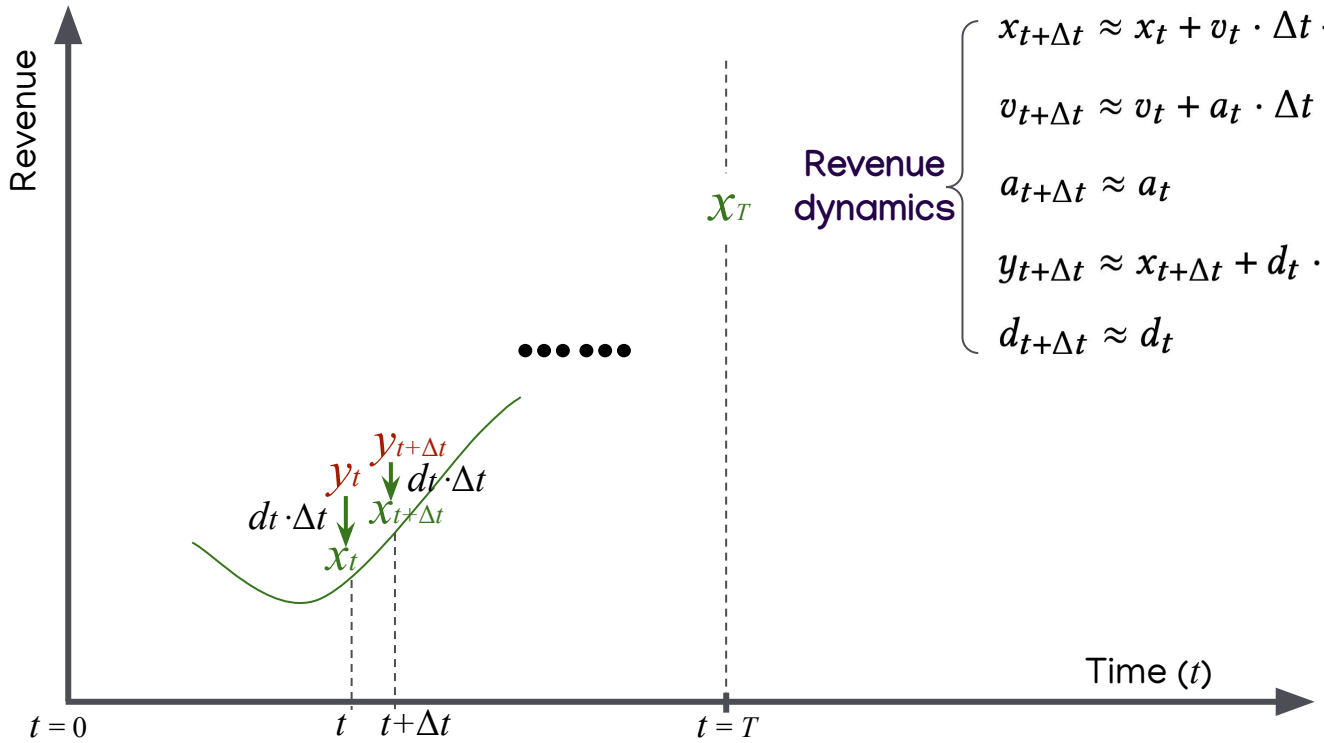
$$a_{t+\Delta t} \approx a_t$$

$$y_{t+\Delta t} \approx x_{t+\Delta t} + d_t \cdot \Delta t$$

The unit error can be largely regarded as "static":

$$d_{t+\Delta t} \approx d_t$$

Revenue Model



Revenue dynamics

$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

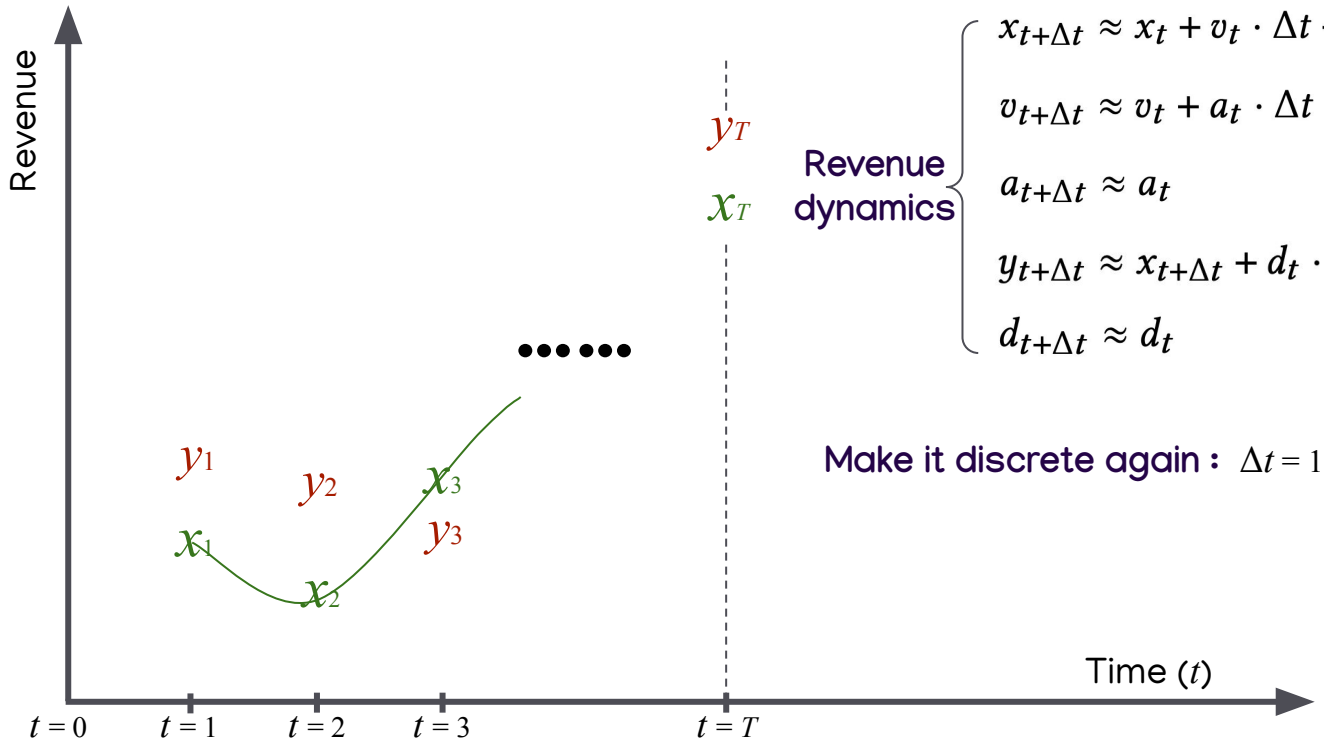
$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

$$a_{t+\Delta t} \approx a_t$$

$$y_{t+\Delta t} \approx x_{t+\Delta t} + d_t \cdot \Delta t$$

$$d_{t+\Delta t} \approx d_t$$

Revenue Model



Revenue dynamics

$$x_{t+\Delta t} \approx x_t + v_t \cdot \Delta t + \frac{1}{2} a_t \cdot \Delta t^2$$

$$v_{t+\Delta t} \approx v_t + a_t \cdot \Delta t$$

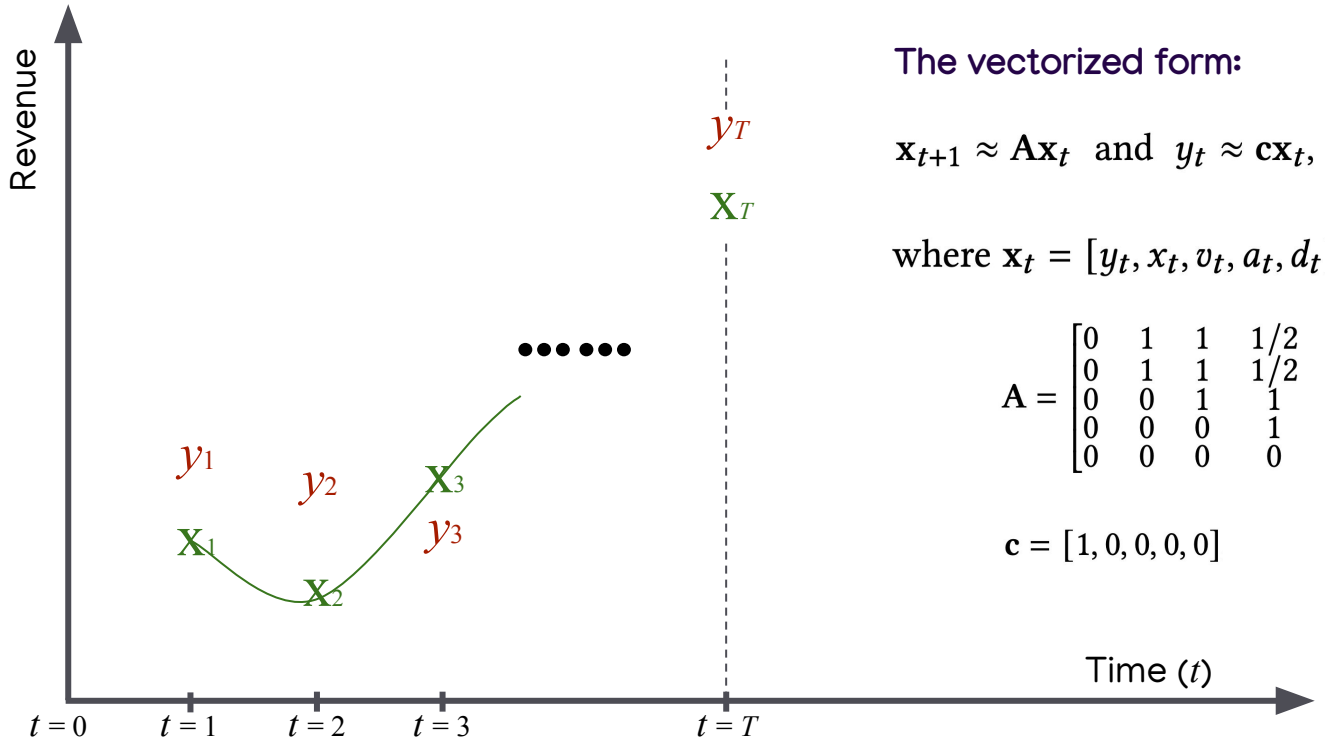
$$a_{t+\Delta t} \approx a_t$$

$$y_{t+\Delta t} \approx x_{t+\Delta t} + d_t \cdot \Delta t$$

$$d_{t+\Delta t} \approx d_t$$

Make it discrete again : $\Delta t = 1$

Revenue Model



The vectorized form:

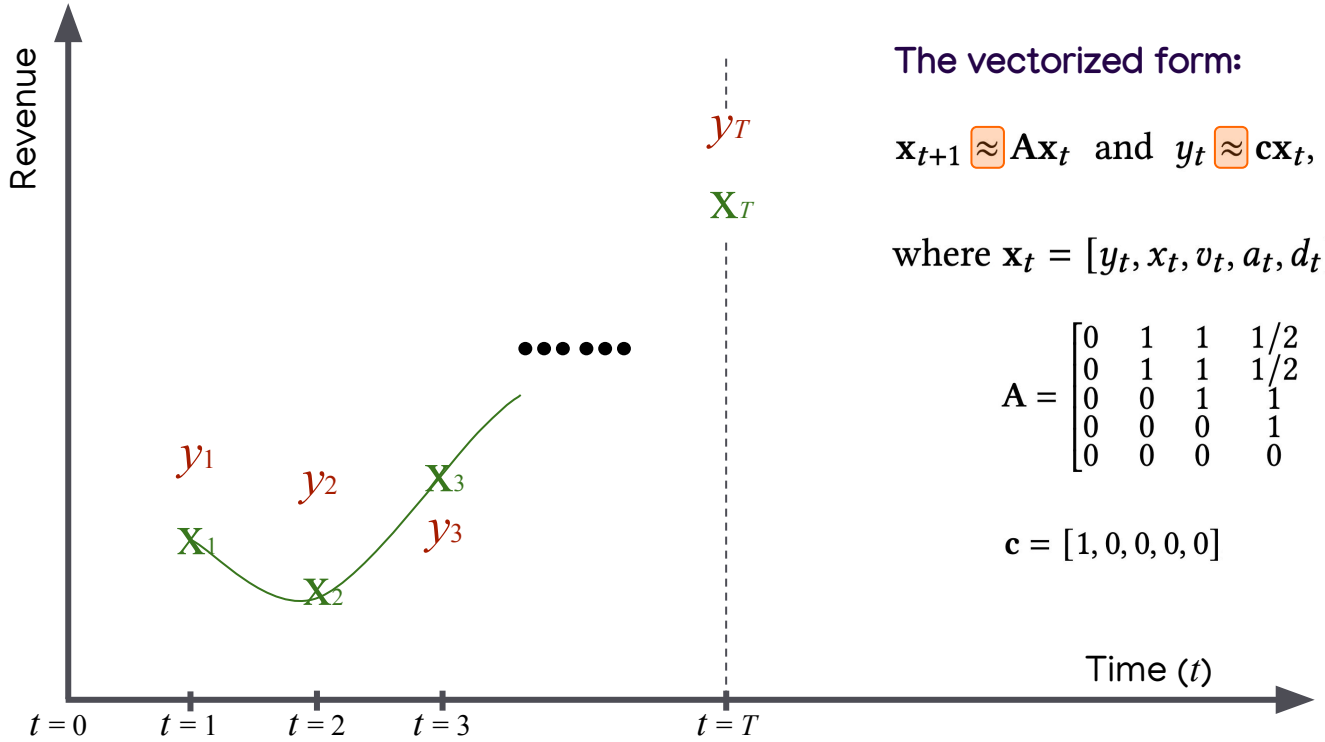
$$\mathbf{x}_{t+1} \approx \mathbf{A}\mathbf{x}_t \quad \text{and} \quad y_t \approx \mathbf{c}\mathbf{x}_t,$$

where $\mathbf{x}_t = [y_t, x_t, v_t, a_t, d_t]^\top$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1/2 & 1 \\ 0 & 1 & 1 & 1/2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{c} = [1, 0, 0, 0, 0]$$

Revenue Model



The vectorized form:

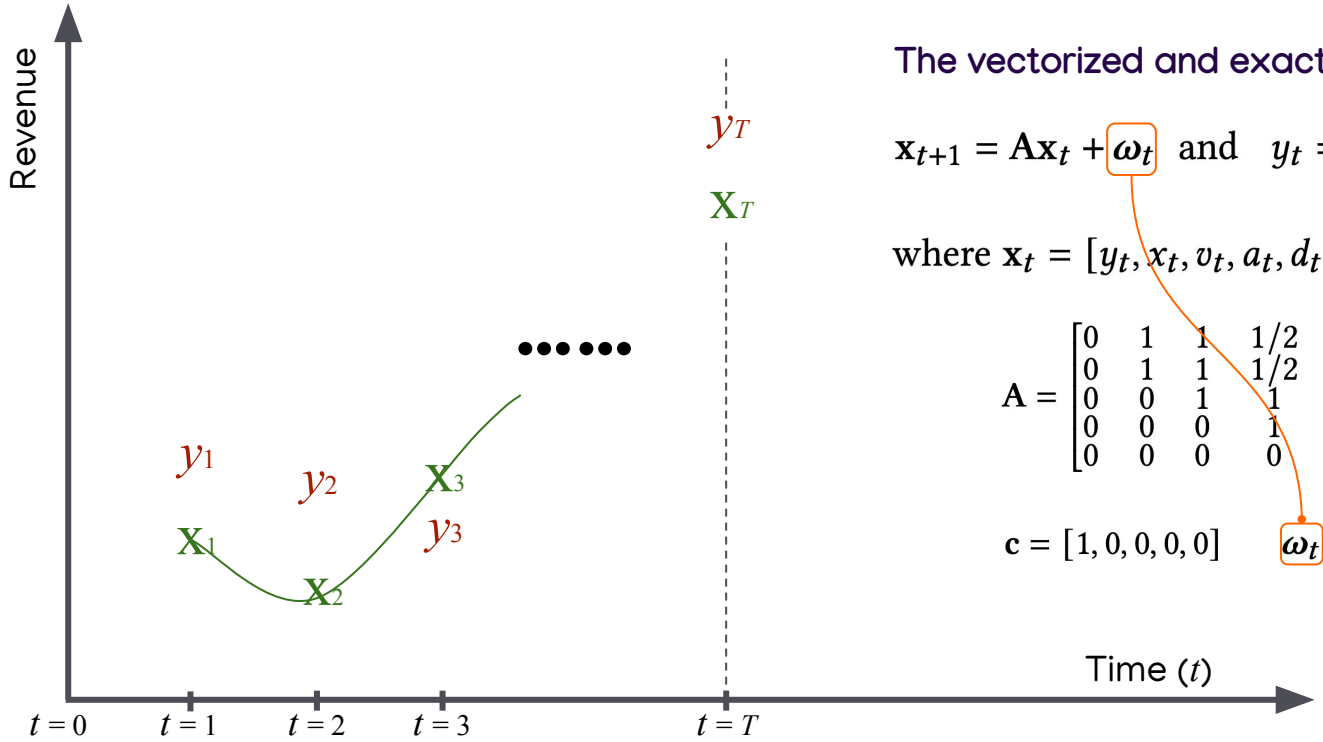
$$\mathbf{x}_{t+1} \approx \mathbf{A}\mathbf{x}_t \quad \text{and} \quad y_t \approx \mathbf{c}\mathbf{x}_t,$$

where $\mathbf{x}_t = [y_t, x_t, v_t, a_t, d_t]^\top$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1/2 & 1 \\ 0 & 1 & 1 & 1/2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{c} = [1, 0, 0, 0, 0]$$

Revenue Model



The vectorized and exact form:

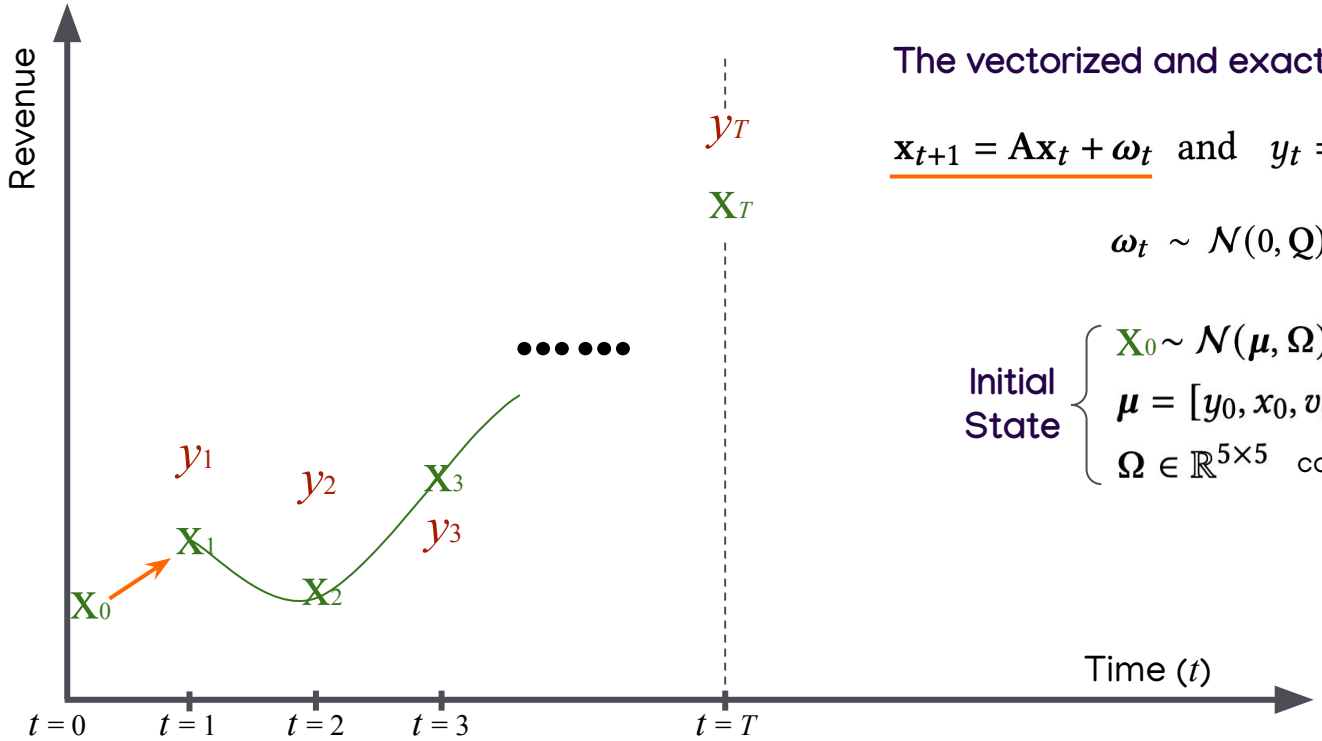
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$$

where $\mathbf{x}_t = [y_t, x_t, v_t, a_t, d_t]^\top$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1/2 & 1 \\ 0 & 1 & 1 & 1/2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{c} = [1, 0, 0, 0, 0] \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \epsilon_t \sim \mathcal{N}(0, \mathbf{R})$$

Revenue Model : Action!



The vectorized and exact form:

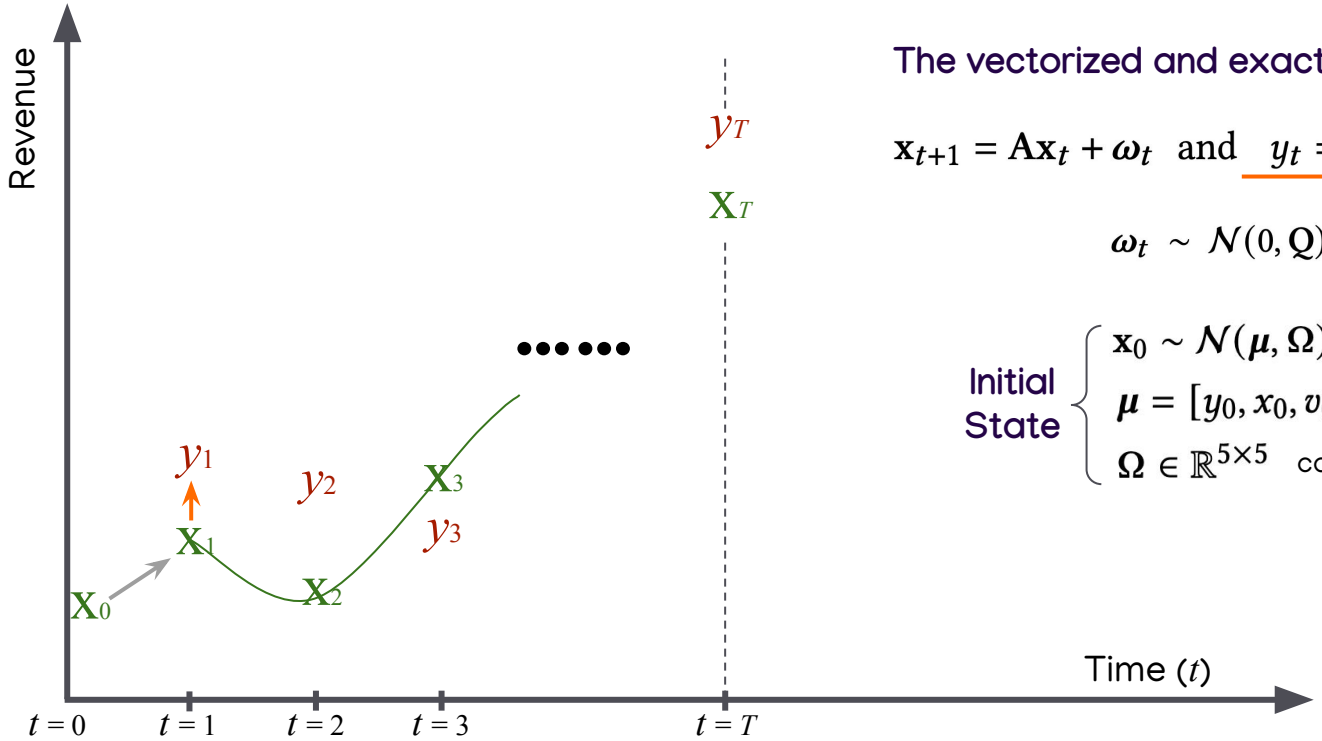
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{X}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Action!



The vectorized and exact form:

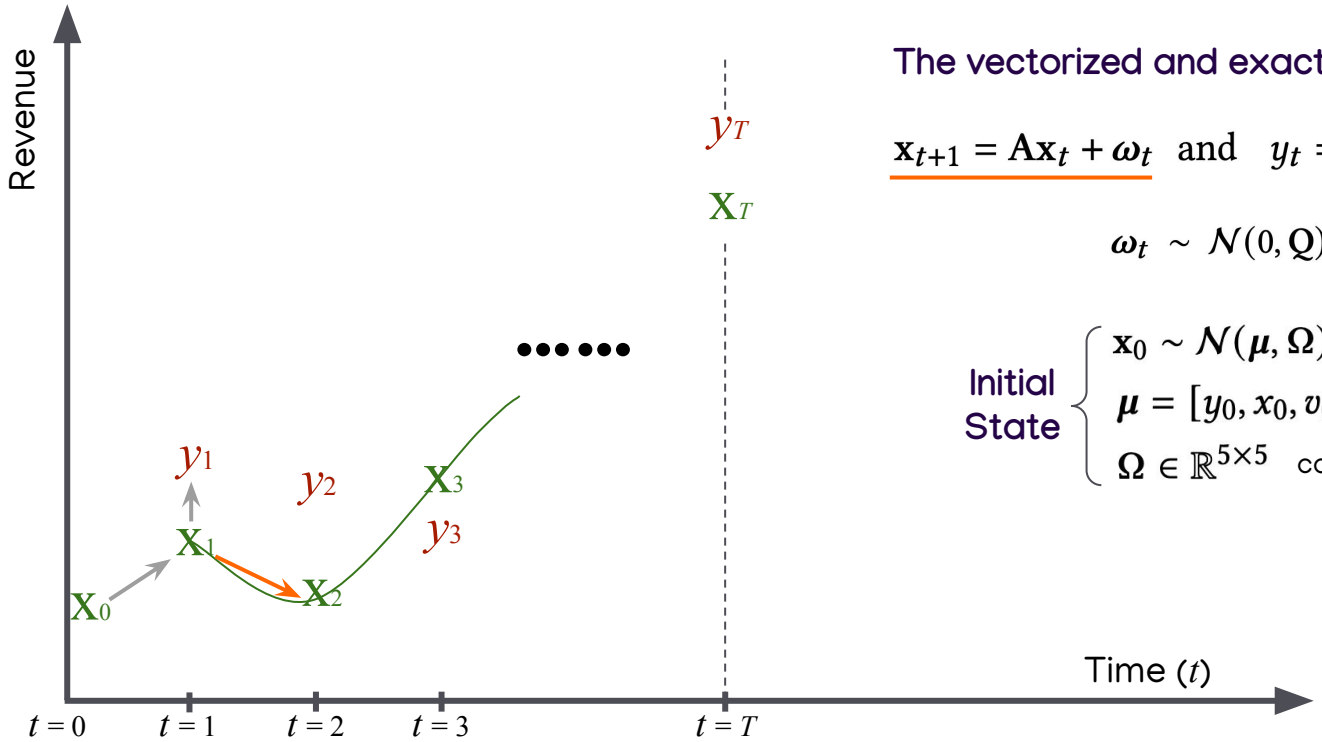
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \epsilon_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Action!



The vectorized and exact form:

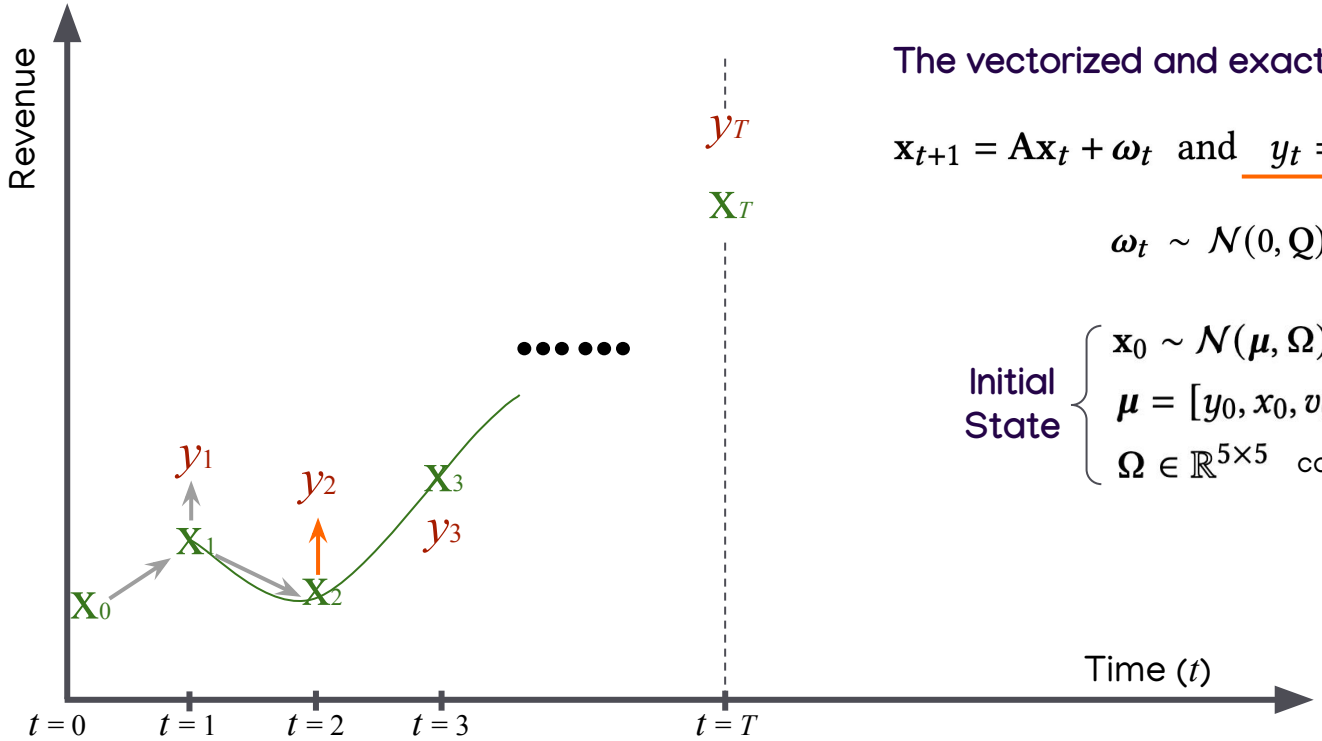
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Action!



The vectorized and exact form:

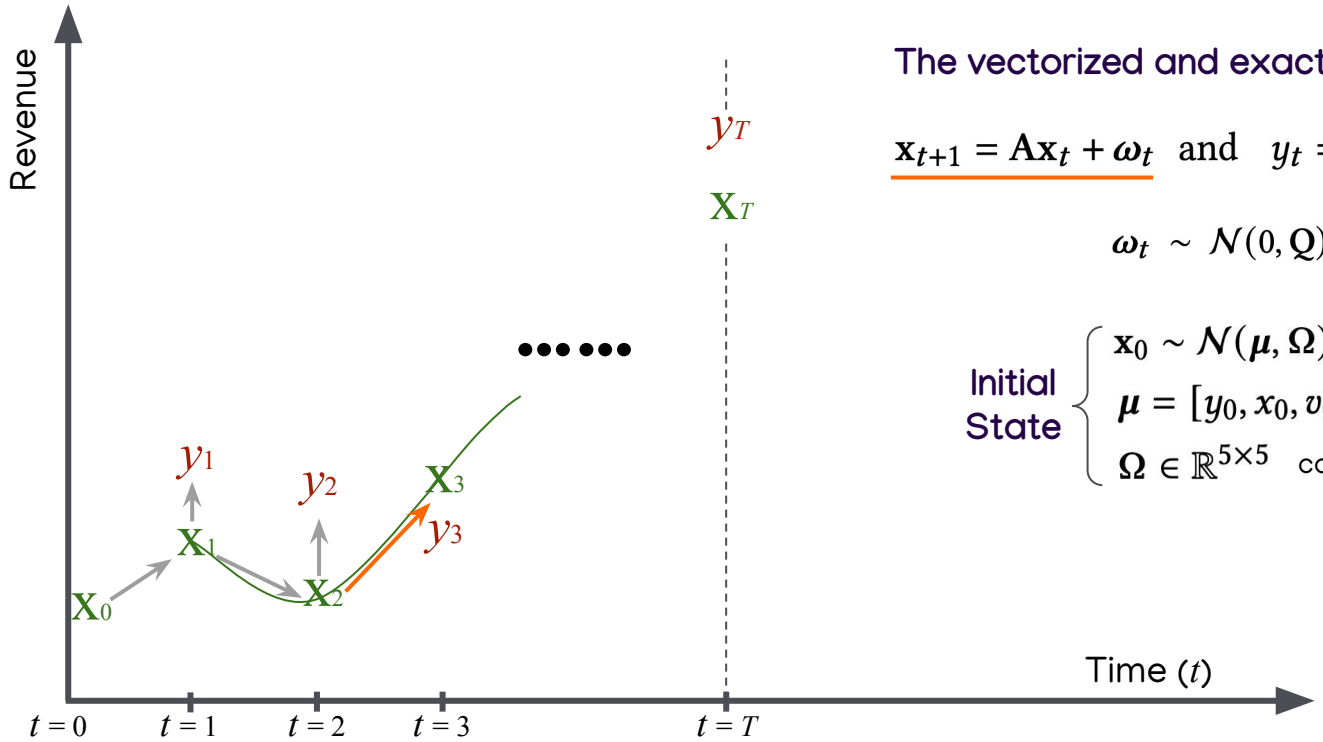
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \epsilon_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Action!



The vectorized and exact form:

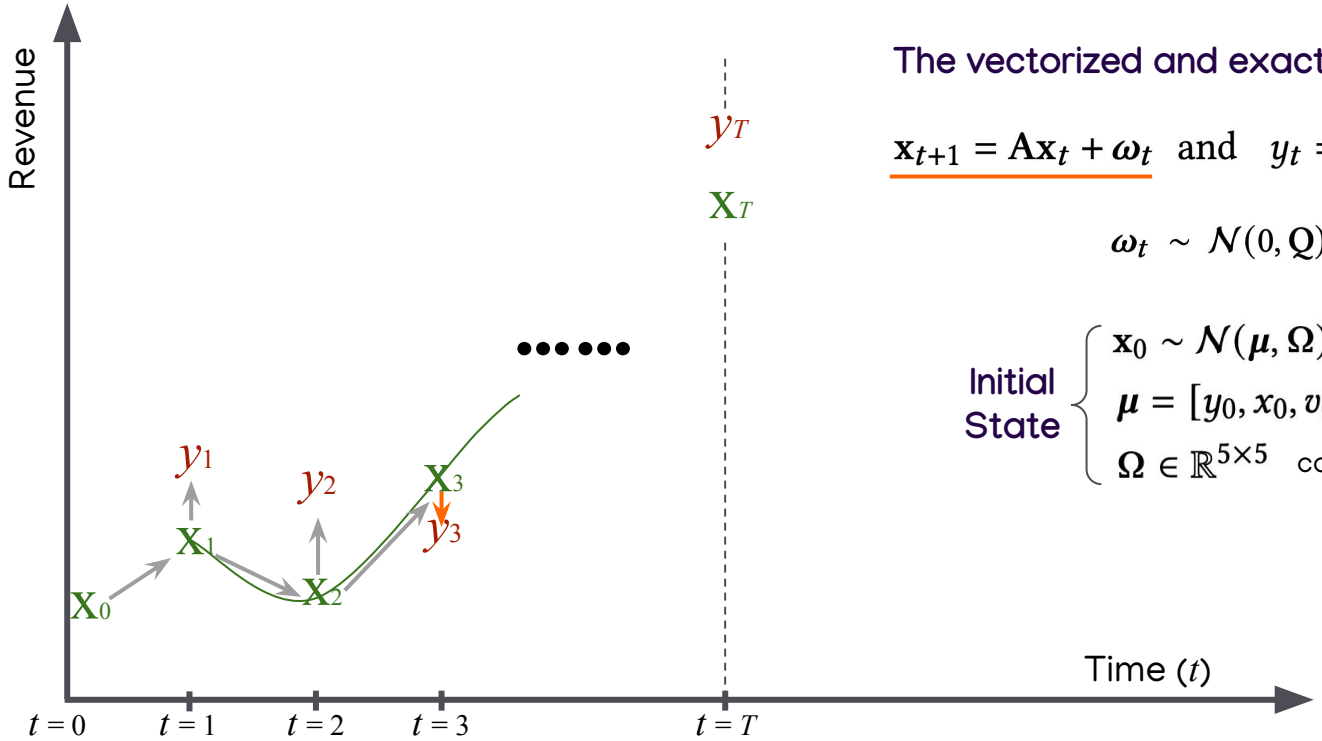
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Action!



The vectorized and exact form:

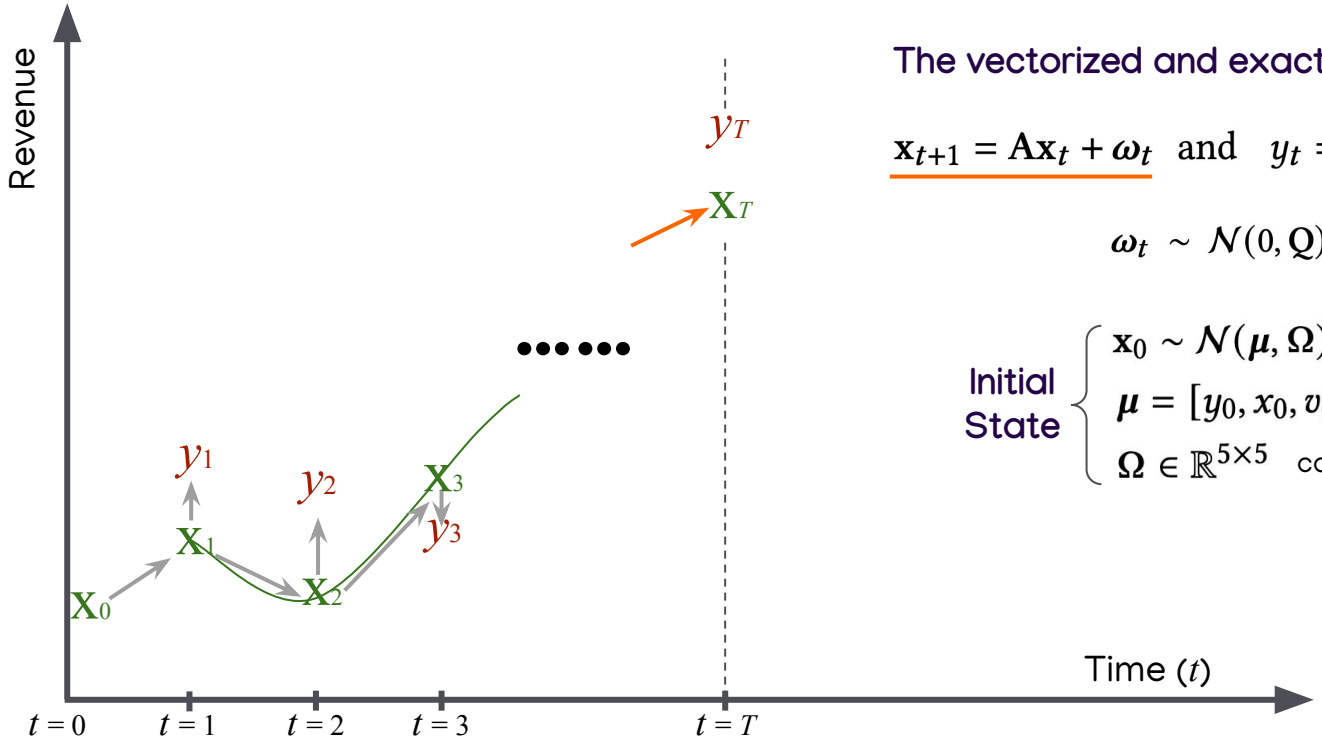
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Action!



The vectorized and exact form:

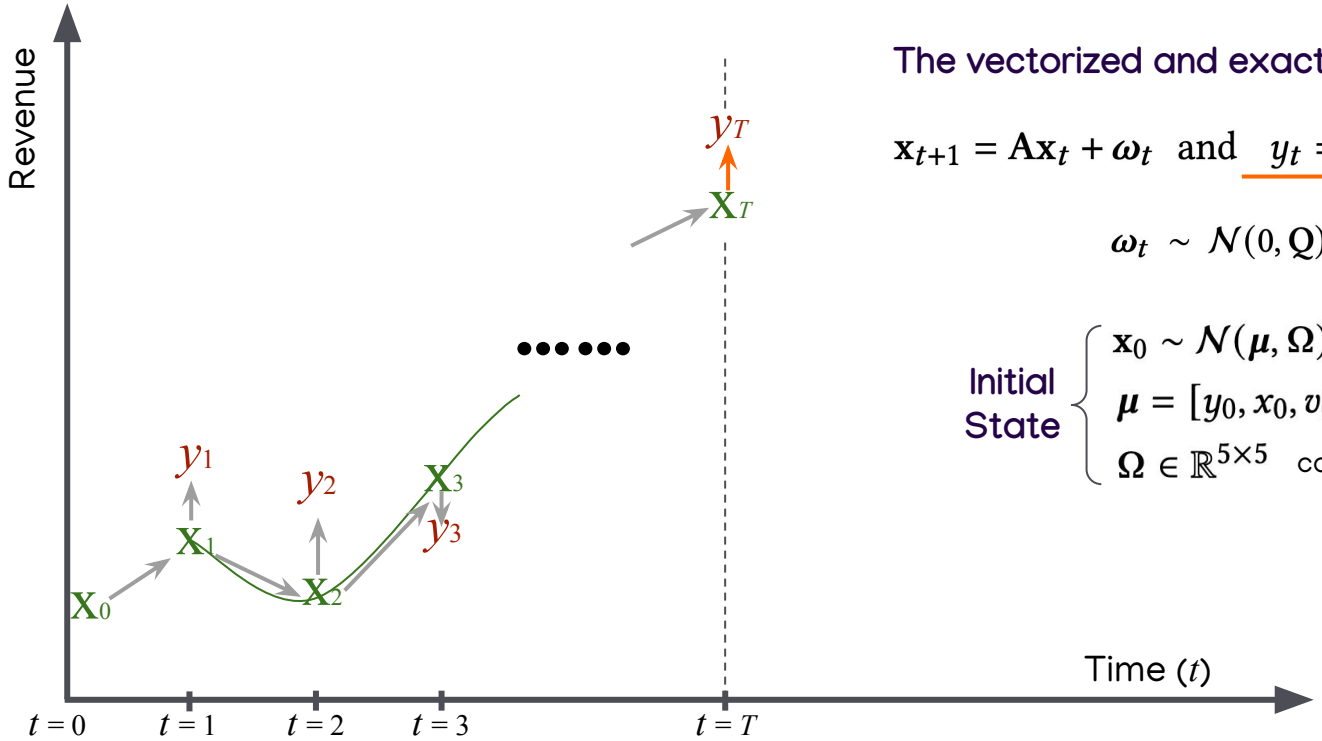
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Action!



The vectorized and exact form:

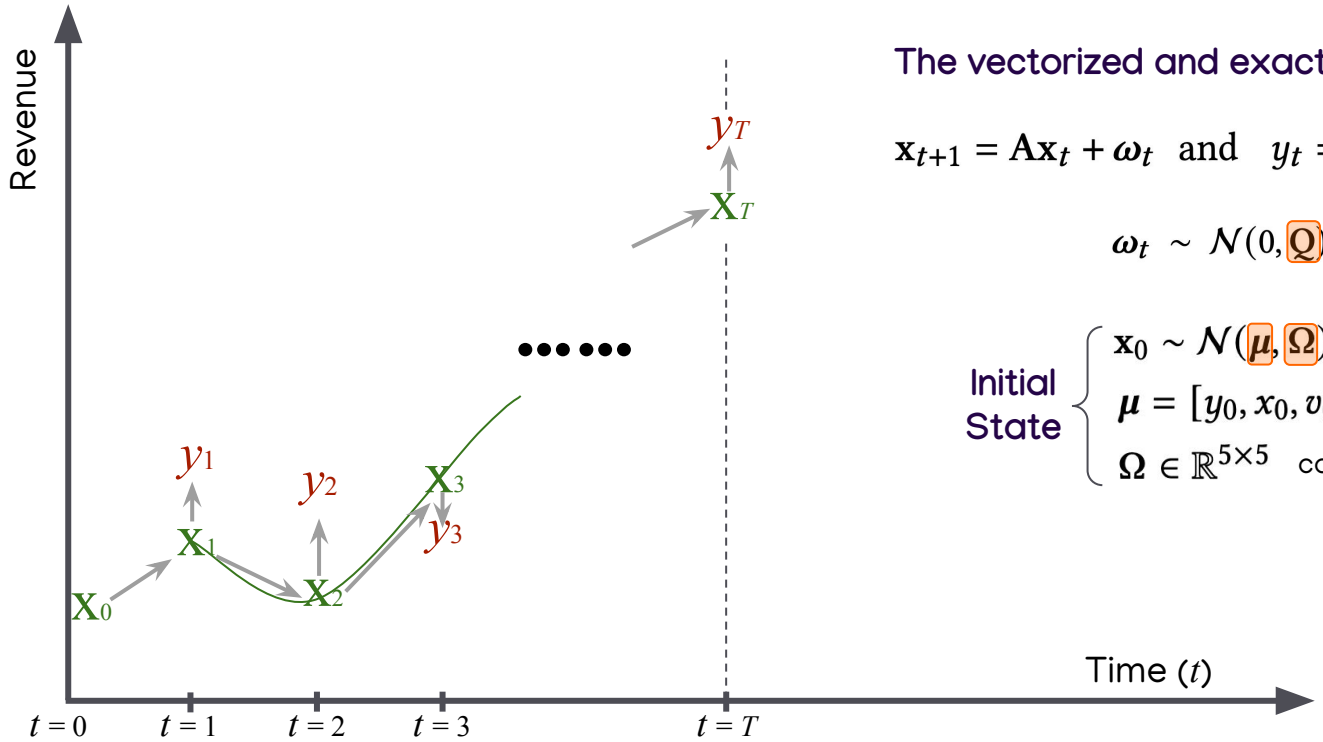
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \epsilon_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Parameters?



The vectorized and exact form:

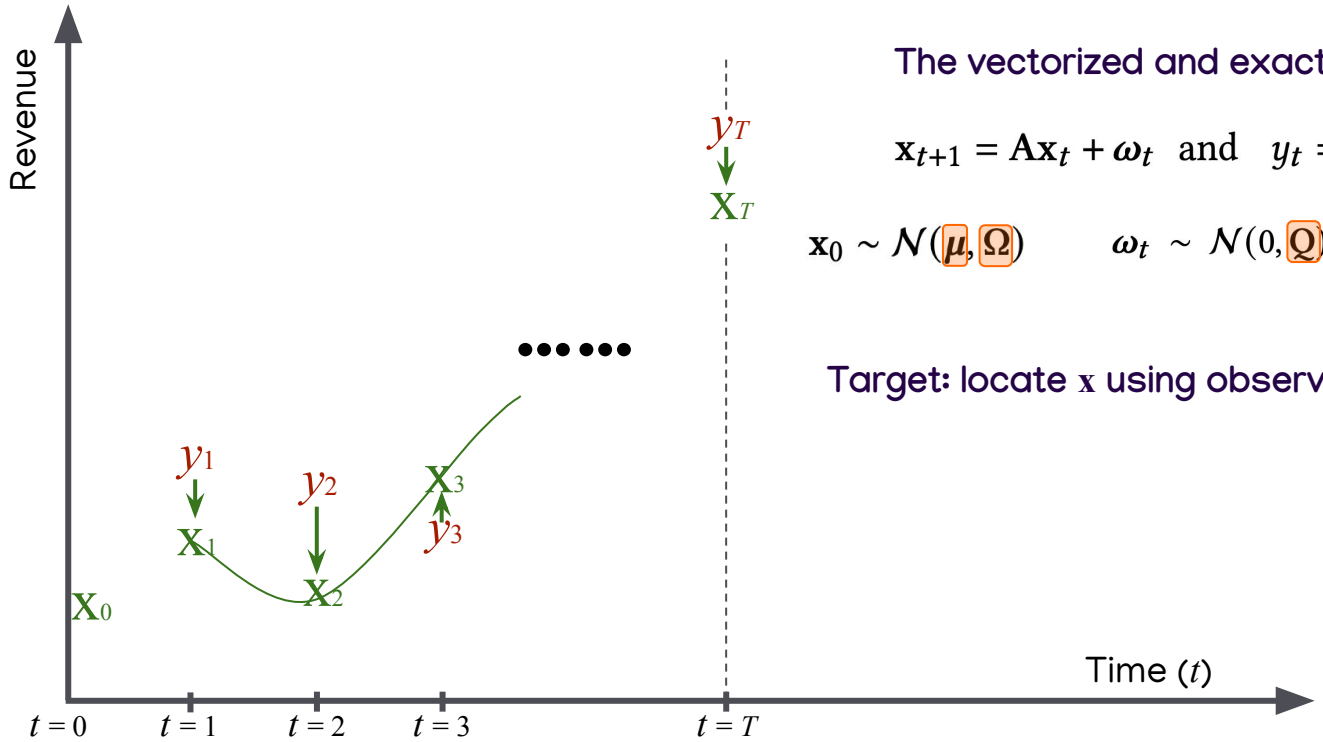
$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{R})$$

Initial State

$$\begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \boldsymbol{\mu} = [y_0, x_0, v_0, a_0, d_0]^\top \\ \boldsymbol{\Omega} \in \mathbb{R}^{5\times 5} \quad \text{covariance matrix} \end{cases}$$

Revenue Model : Optimization!



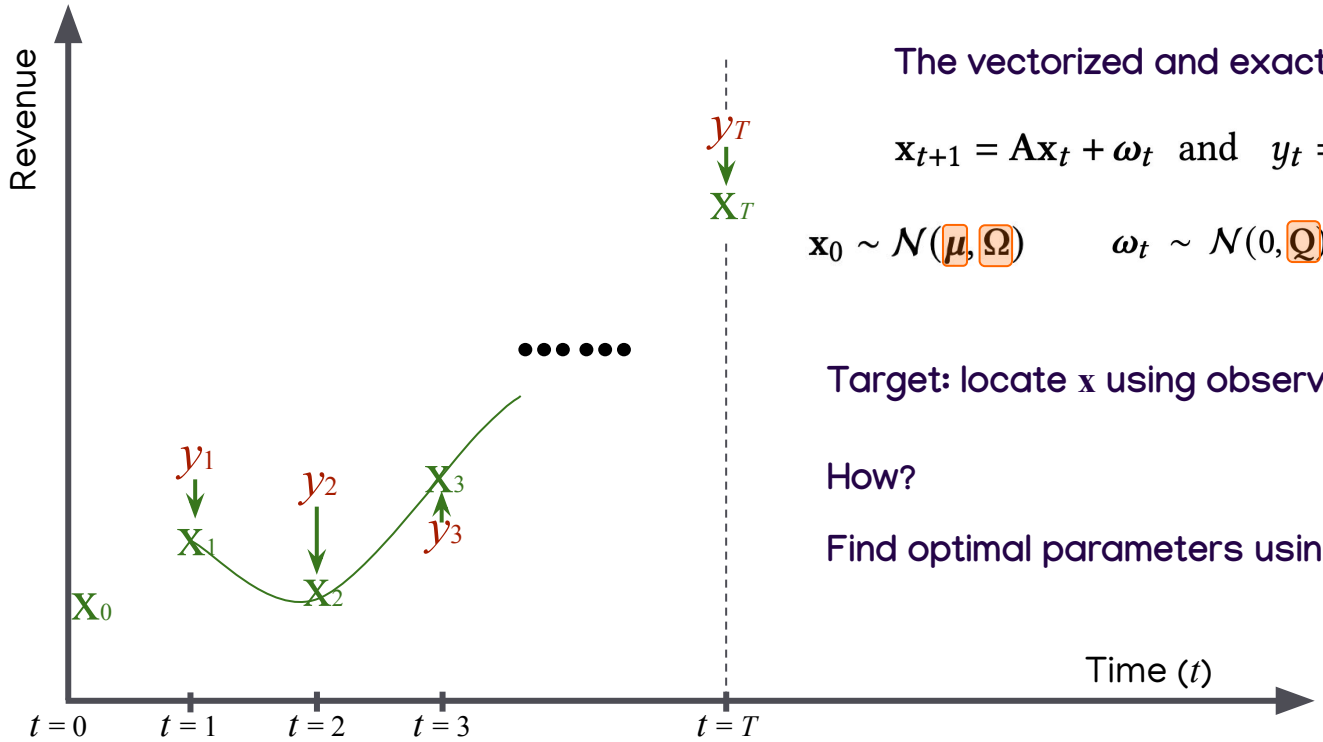
The vectorized and exact form:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$$

$$\mathbf{x}_0 \sim \mathcal{N}(\mu, \Omega) \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathrm{Q}) \qquad \epsilon_t \sim \mathcal{N}(0, \mathrm{R})$$

Target: locate $\mathbf{x}$ using observed measurement y

Revenue Model : Optimization!



The vectorized and exact form:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$$

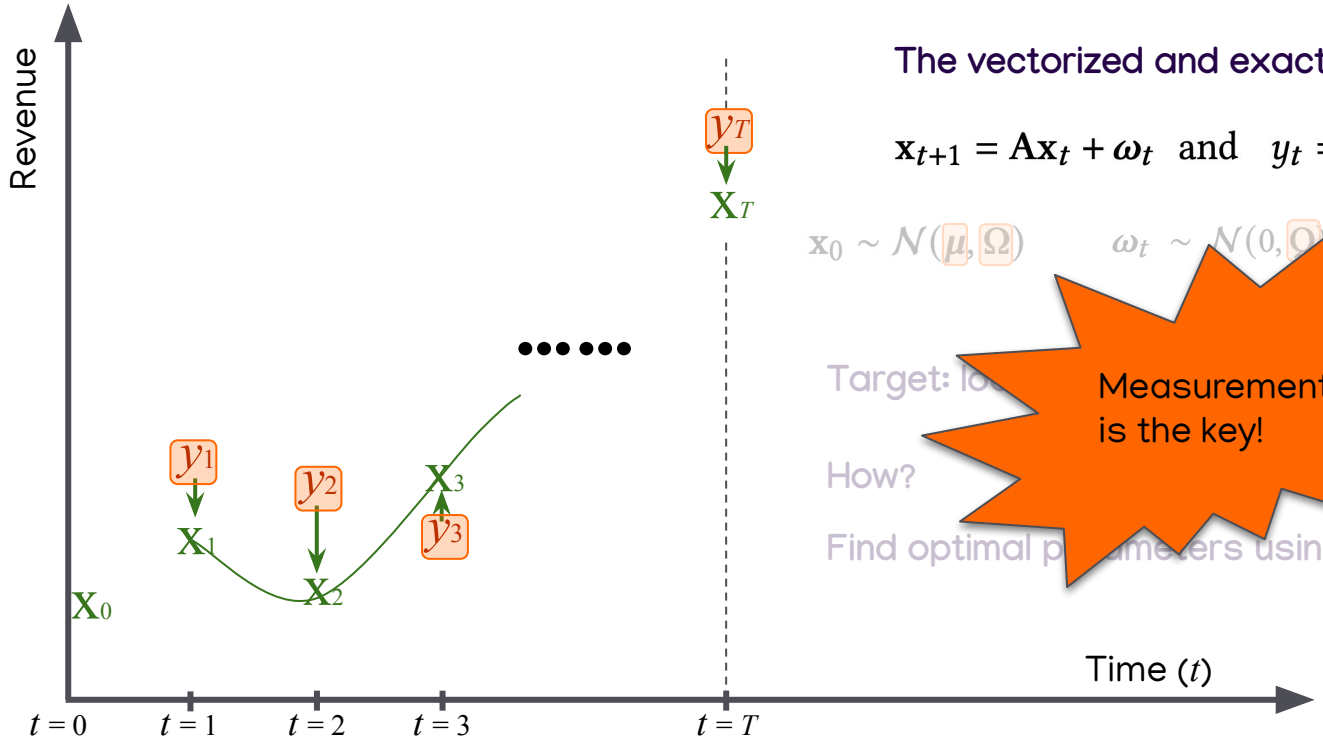
$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}, \Omega) \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathrm{Q}) \qquad \epsilon_t \sim \mathcal{N}(0, \mathrm{R})$$

Target: locate $\mathbf{x}$ using observed measurement y

How?

Find optimal parameters using EM algorithm!

Revenue Model : Optimization!



The vectorized and exact form:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t \quad \text{and} \quad y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$$

$\mathbf{x}_0 \sim \mathcal{N}(\mu, \Omega)$ $\omega_t \sim \mathcal{N}(0, \Omega)$ $\epsilon_t \sim \mathcal{N}(0, \mathrm{R})$

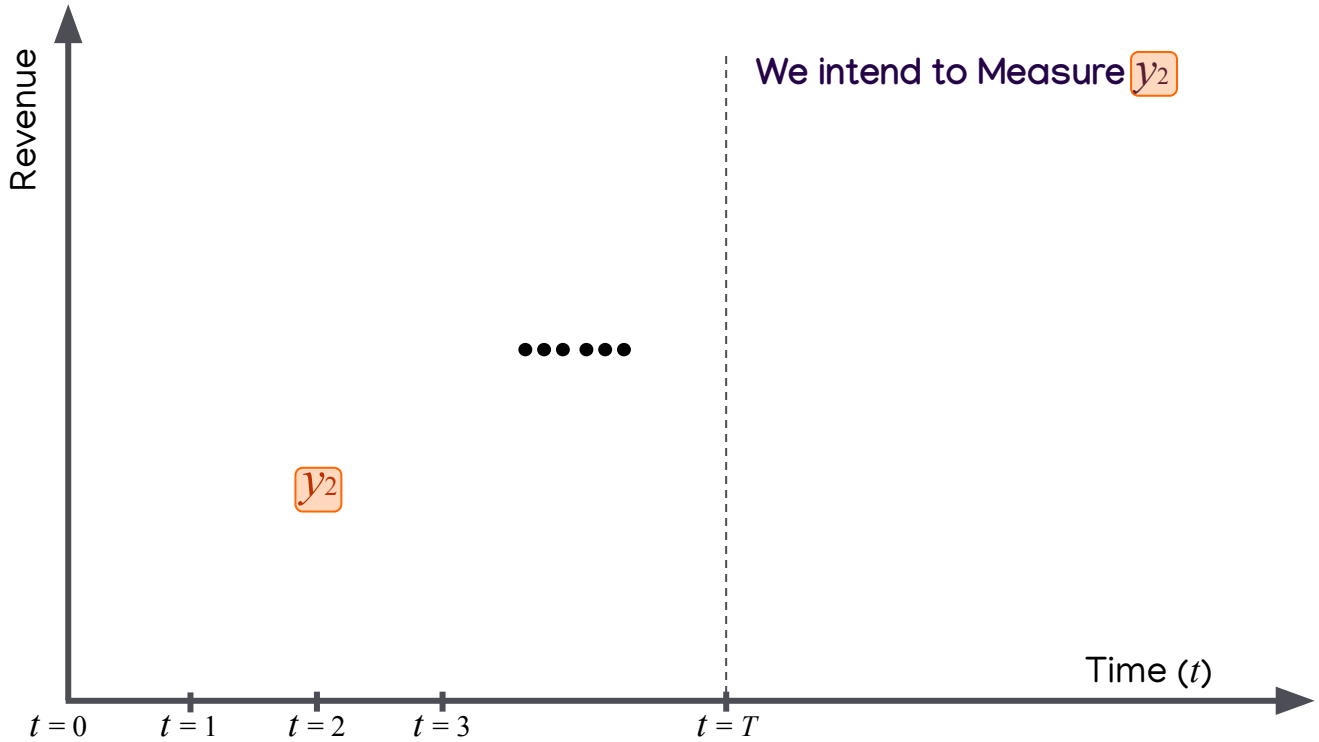
Target: lo… …surement y

How?

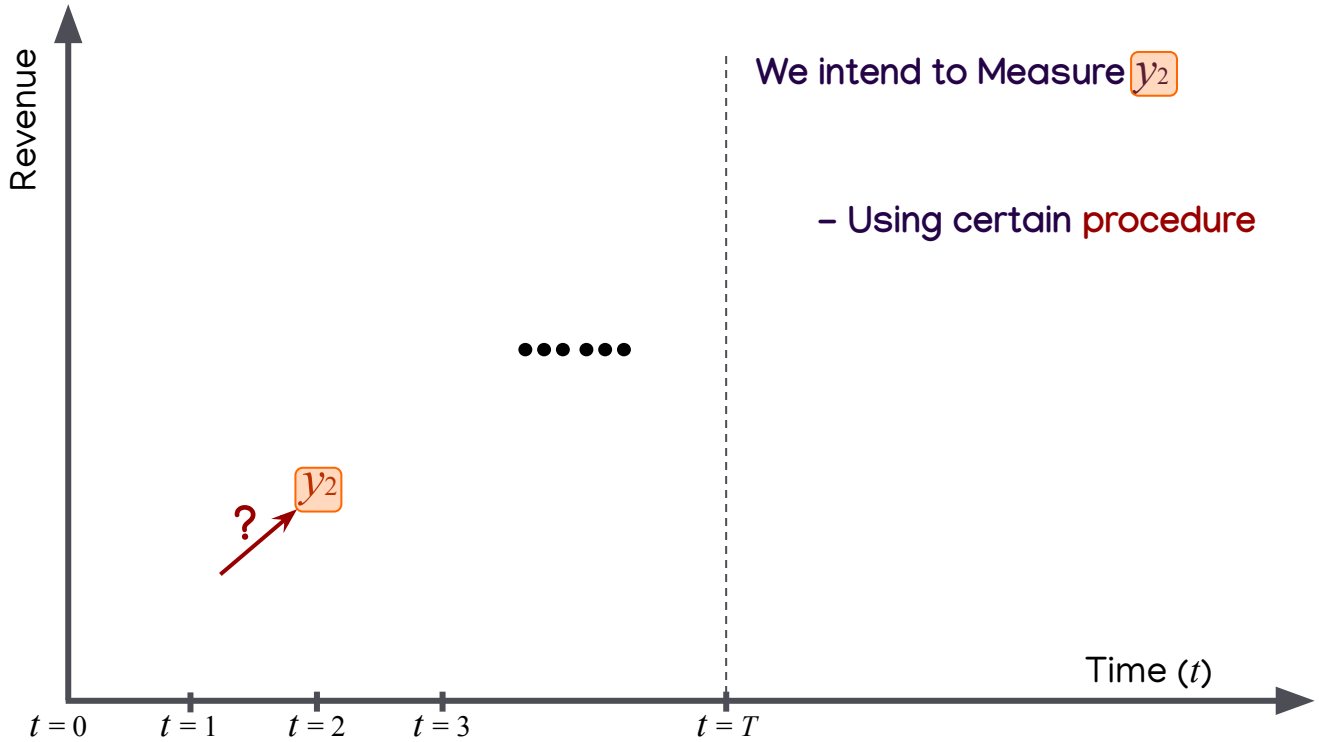
Find optimal p…meters using EM algorithm!

Measurement is the key!

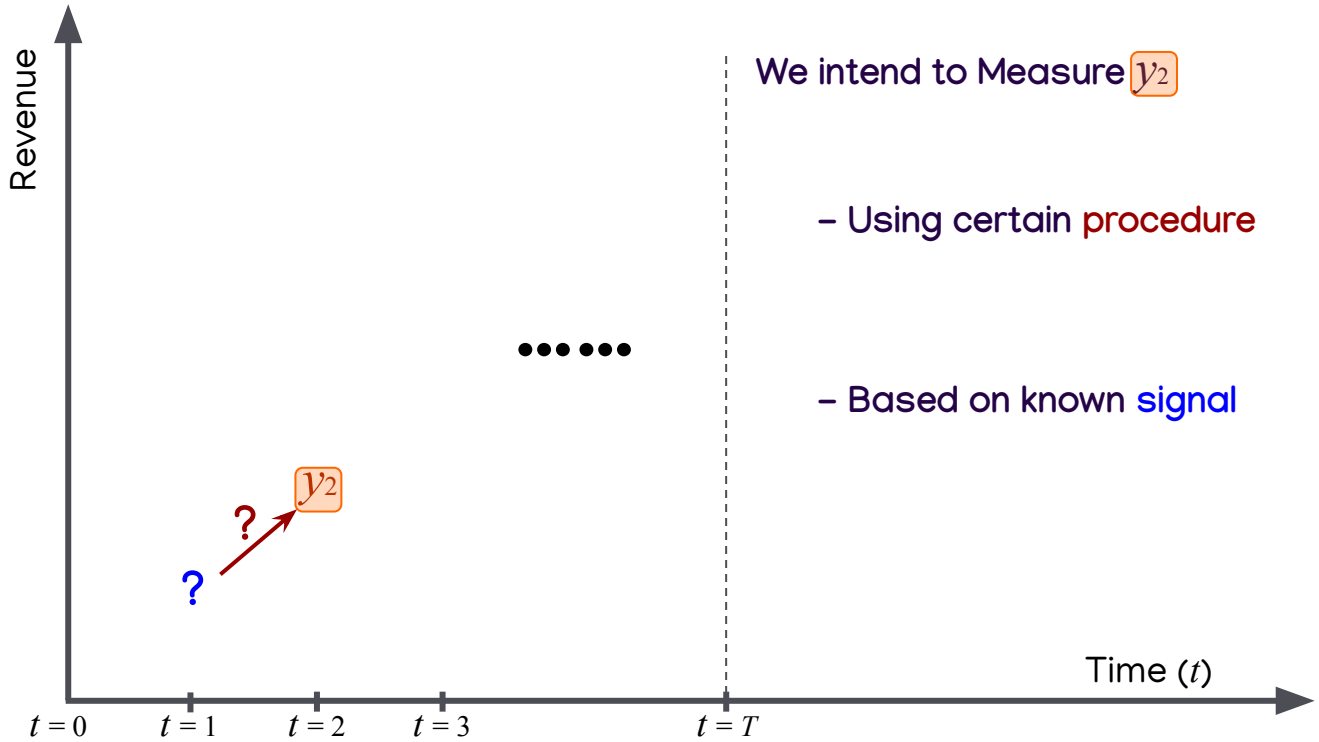
Measurements



Measurements



Measurements

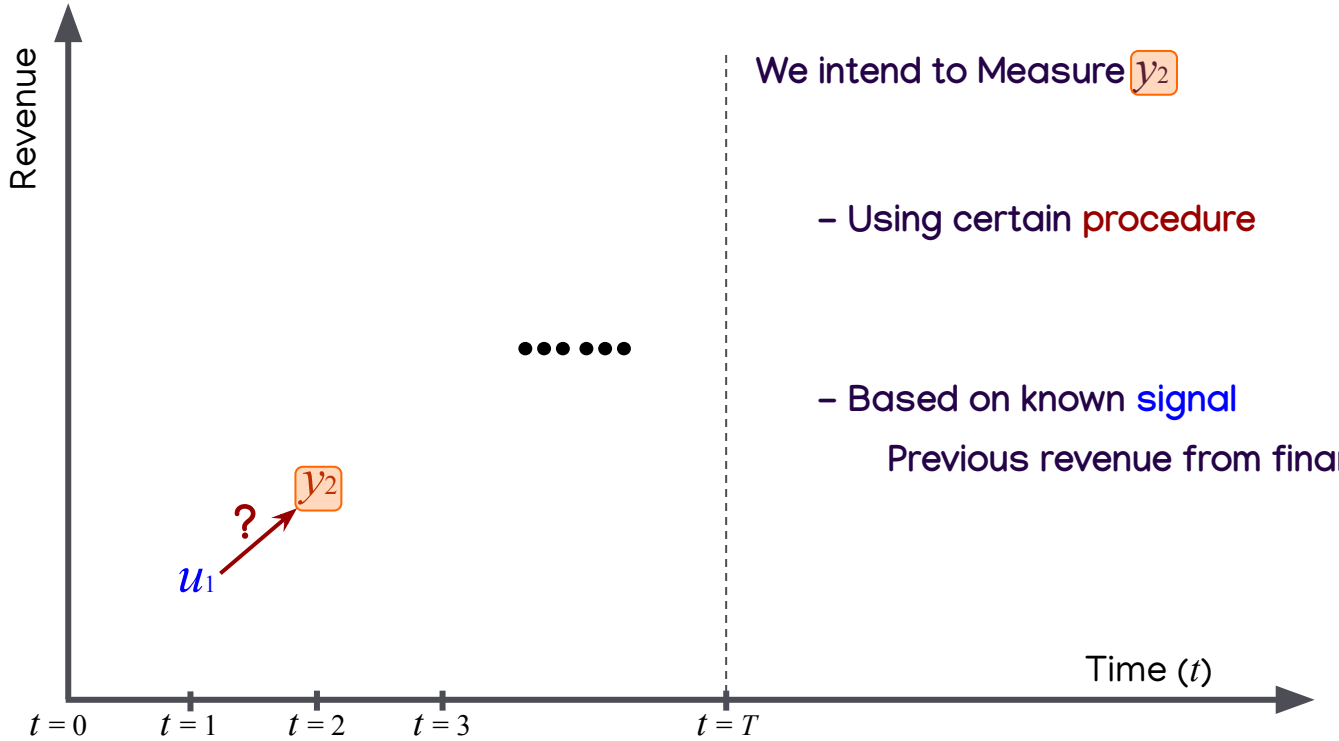


We intend to Measure y_2

– Using certain procedure

– Based on known signal

Measurements



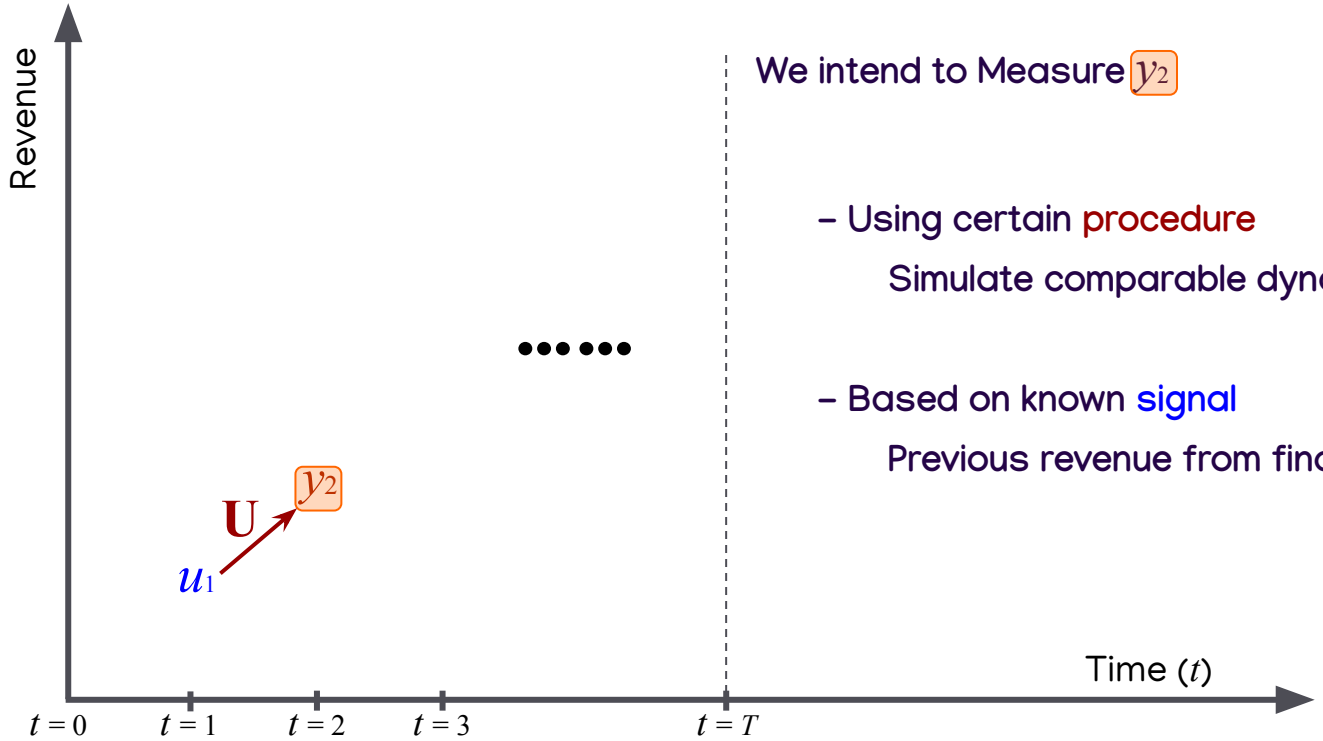
We intend to Measure y_2

– Using certain procedure

– Based on known signal

Previous revenue from financial report u_1

Measurements



We intend to Measure y_2

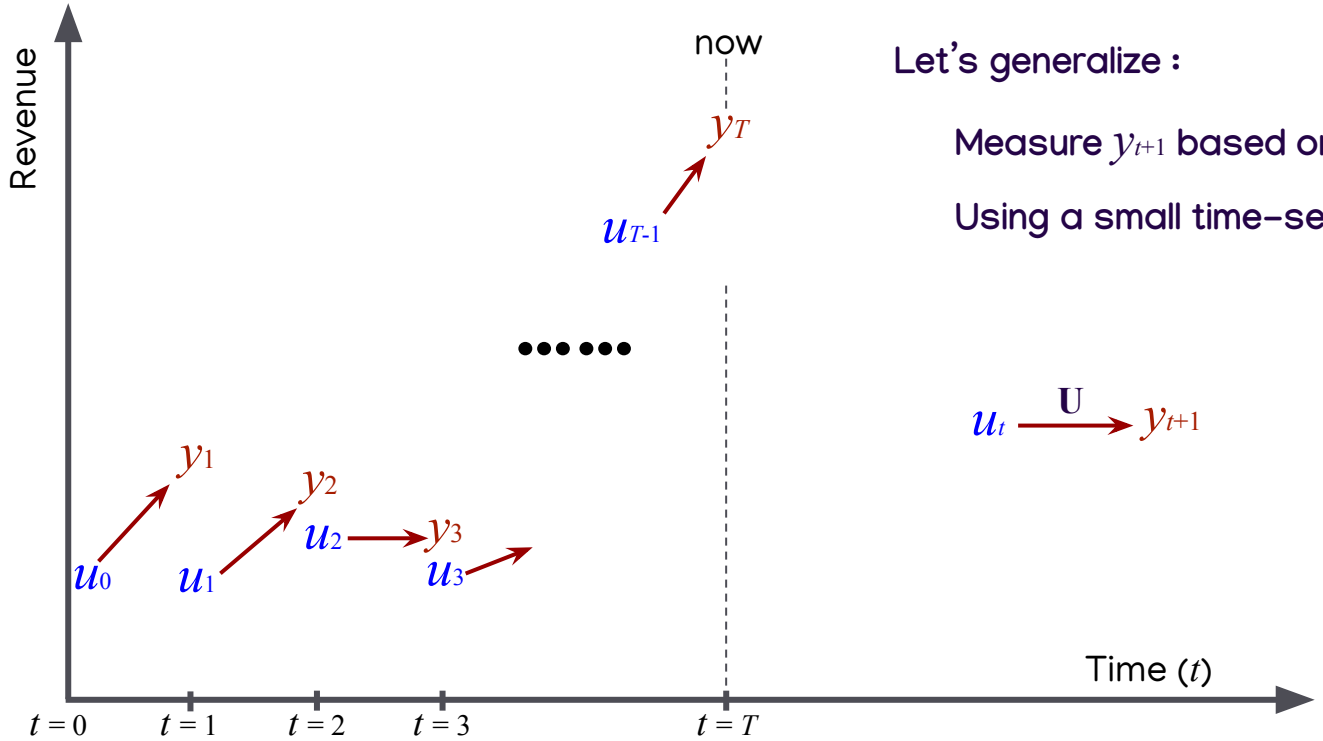
– Using certain procedure

Simulate comparable dynamics in a dataset **U**

– Based on known signal

Previous revenue from financial report u_1

Measurements



Let's generalize :

Measure y_{t+1} based on u_t

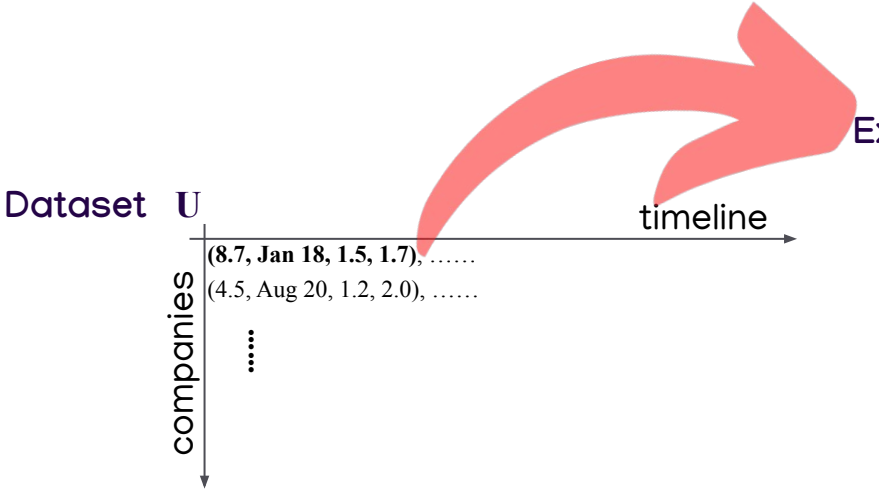
Using a small time-series dataset $\mathbf{U}$

$$u_t \xrightarrow{\mathbf{U}} y_{t+1}$$

Measurements

Measurements

$u_t \xrightarrow{\mathbf{U}} y_{t+1}$



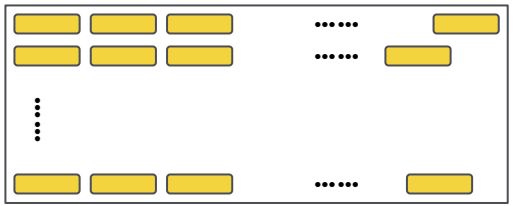
Explain a Tuple **(8.7, Jan 18, 1.5, 1.7)**

- **Jan 18** – the time when obtaining this tuple
- **8.7** – the revenue obtained in **Jan 18** is **8.7**
- **1.5** (current YoY growth)
  – the revenue of **Jan 17** is **8.7/1.5=5.8**
- **1.7** (next YoY growth)
  – the revenue ratio: **Feb 18 / Feb 17 = 1.7**

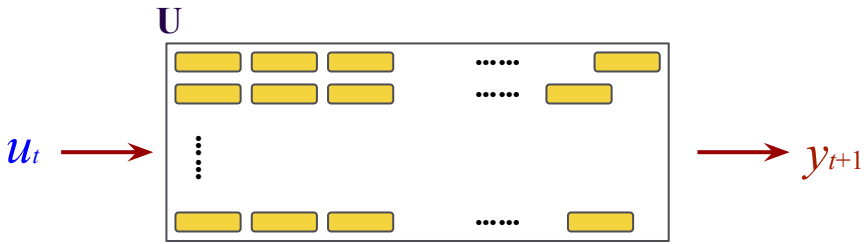
Measurements

$u_t \xrightarrow{\mathbf{U}} y_{t+1}$

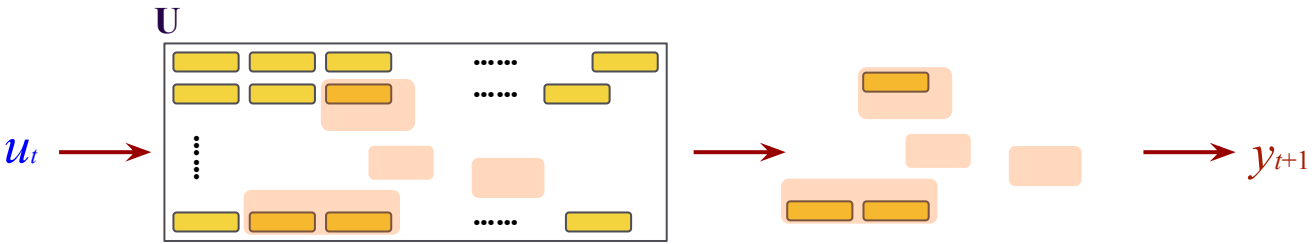
Dataset $\mathbf{U}$



Measurements



Measurements

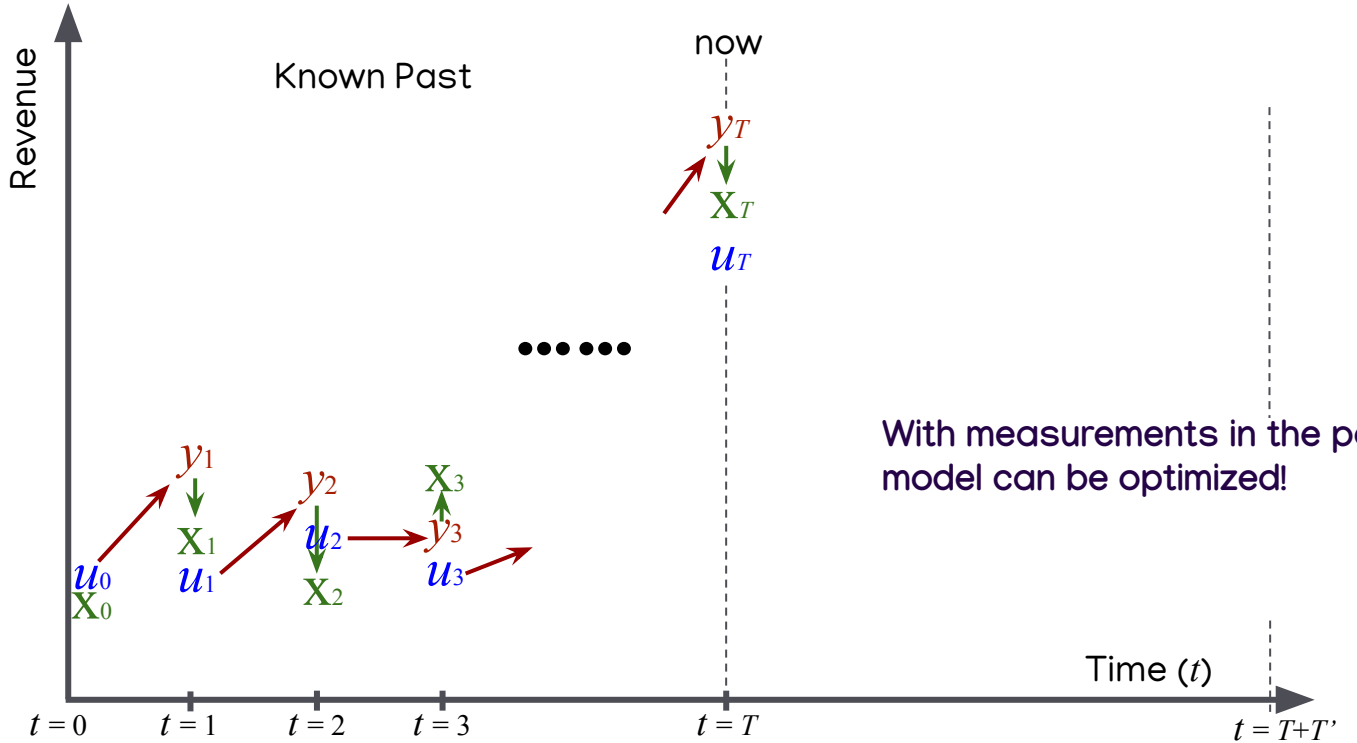


Filtering by :

- business,
- year-month,
- revenue,
- YoY revenue growth

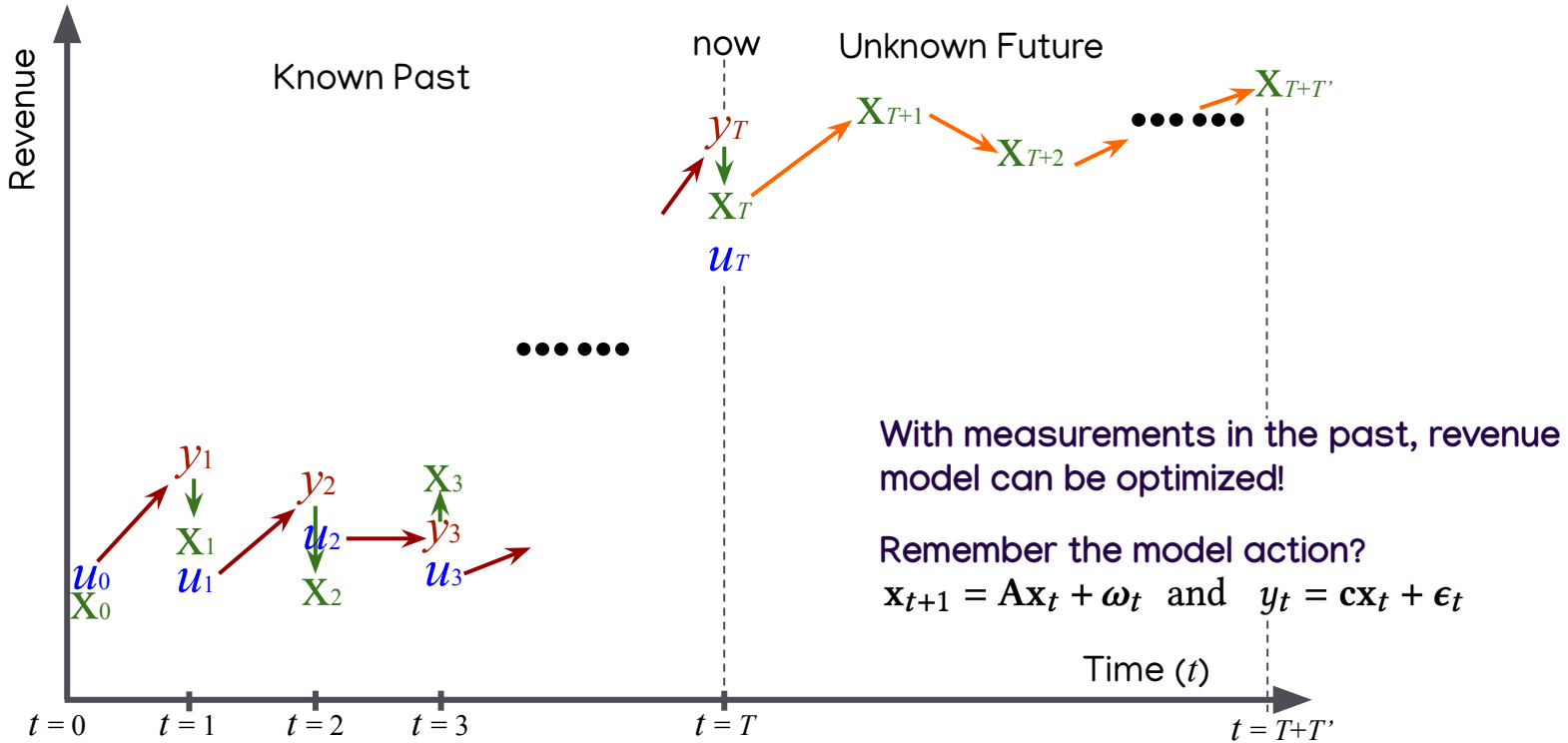
Sampling:
A **stochastic** approach

Measurements

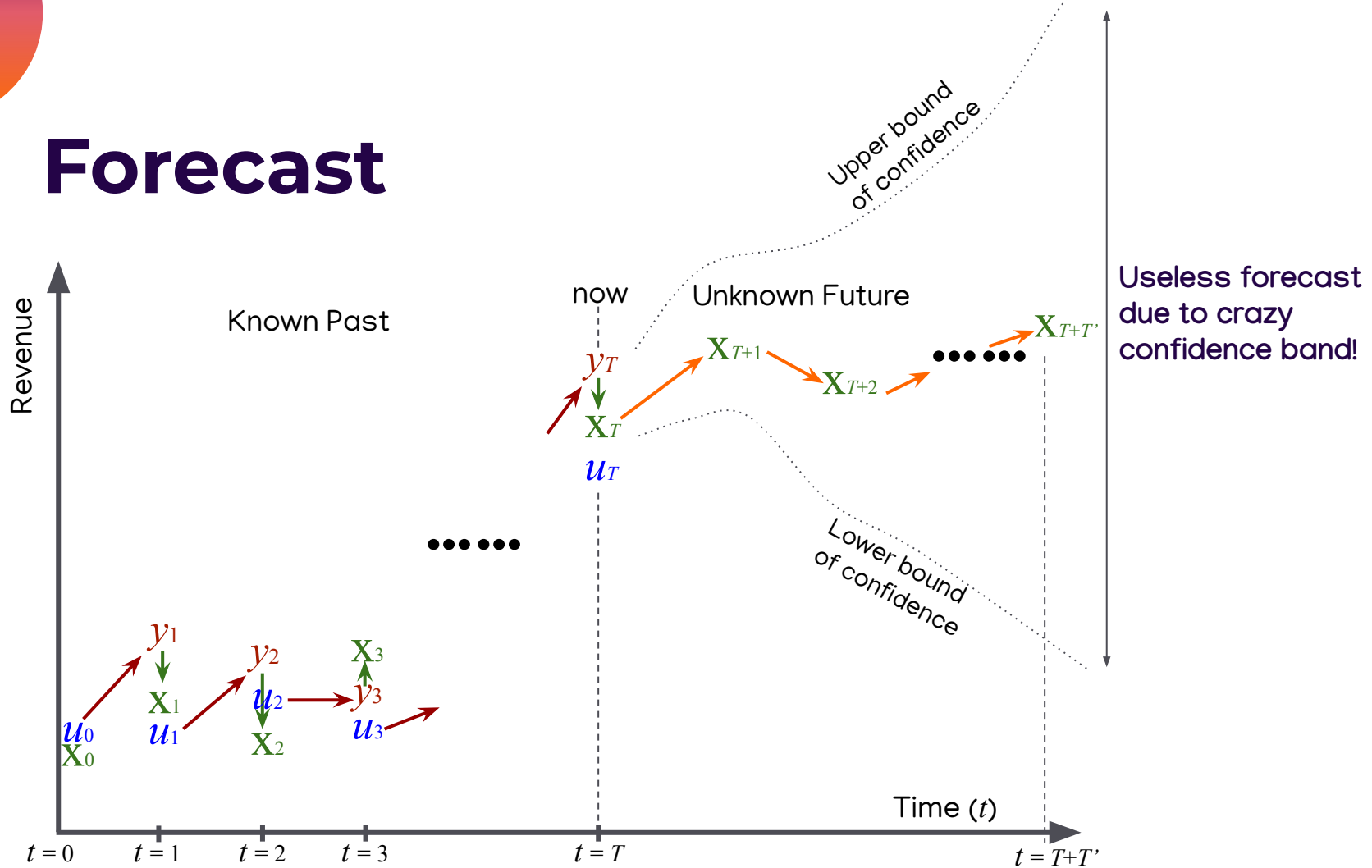


With measurements in the past, revenue model can be optimized!

Forecast

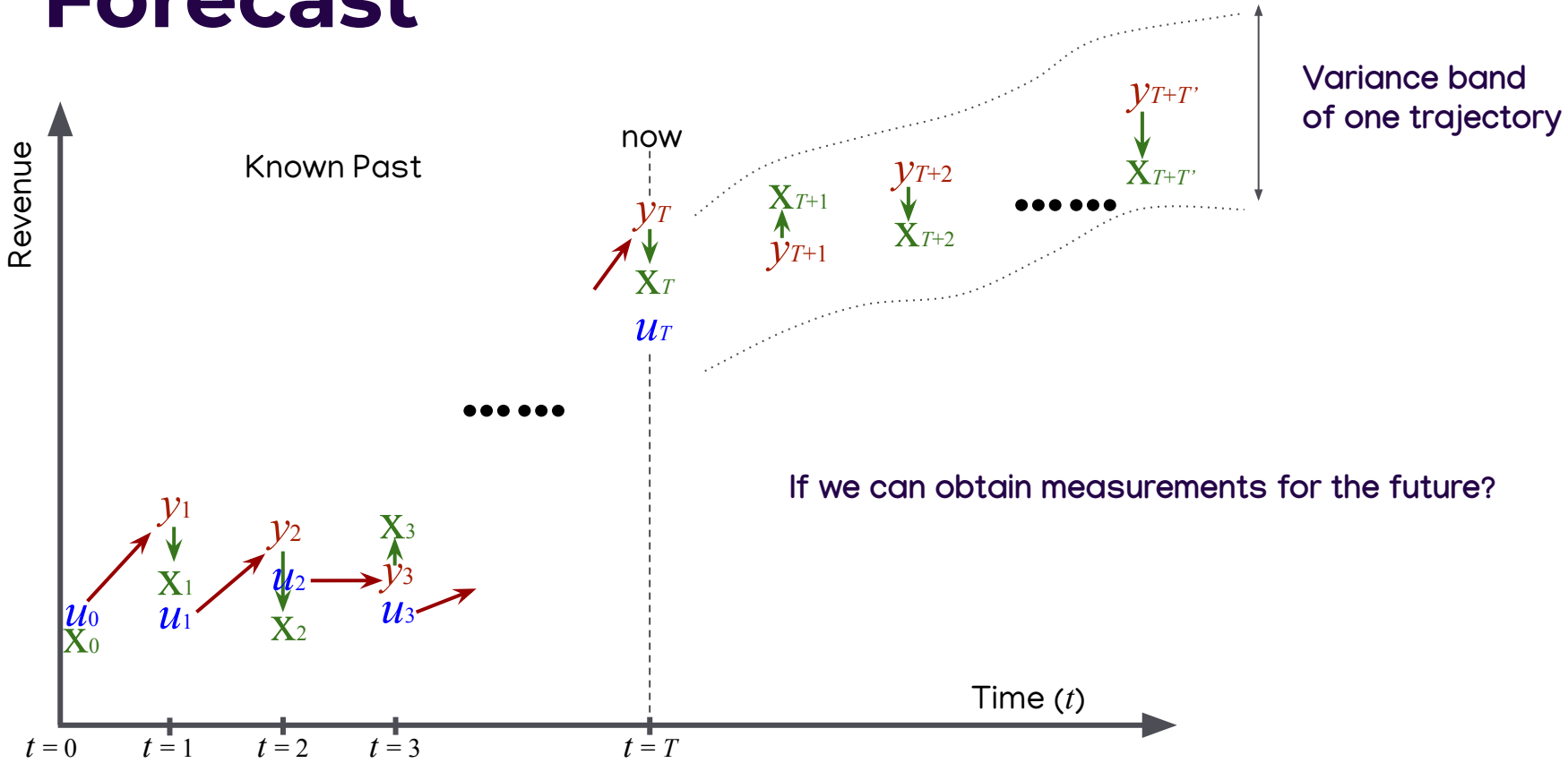

Revenue
Known Past
now
Unknown Future
y_T
X_T
u_T
X_{T+1}
X_{T+2}
$X_{T+T'}$
y_1
X_1
u_1
y_2
u_2
X_2
X_3
y_3
u_3
u_0
X_0
With measurements in the past, revenue model can be optimized!
Remember the model action?
$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\omega}_t$ and $y_t = \mathbf{c}\mathbf{x}_t + \epsilon_t$
Time (t)
$t = 0$
$t = 1$
$t = 2$
$t = 3$
$t = T$
$t = T+T'$

Forecast



Useless forecast due to crazy confidence band!

Forecast



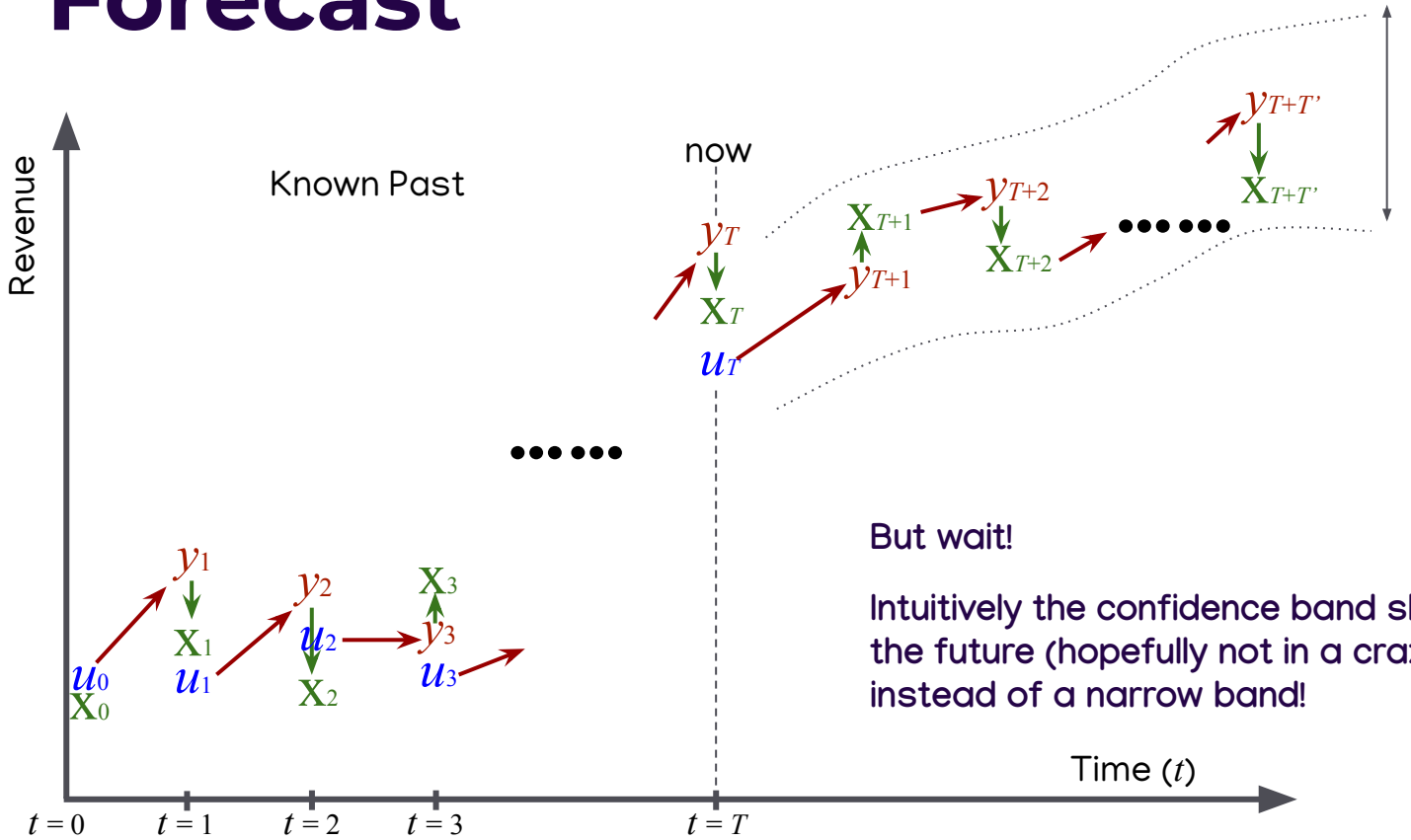
If we can obtain measurements for the future?

Forecast

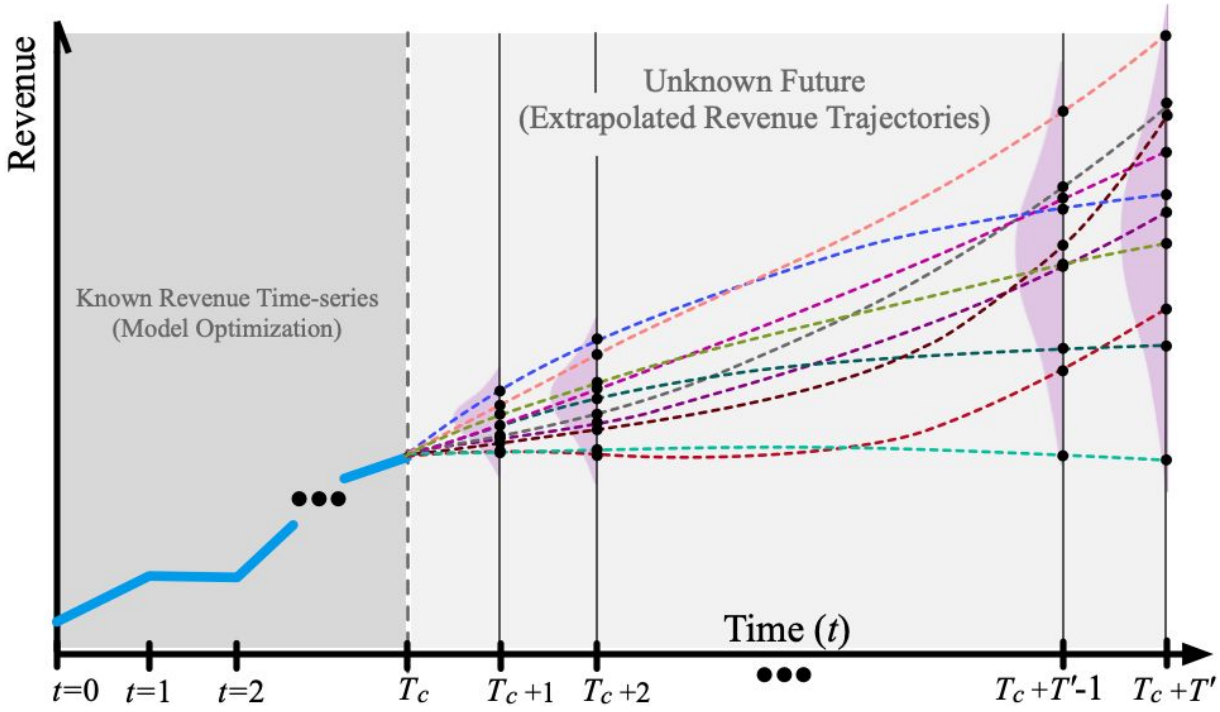
If we can obtain measurements for the future?

No u? Try to base on x instead!

Forecast


Revenue
Known Past
now
Variance band
of one trajectory
But wait!
Intuitively the confidence band should grow into the future (hopefully not in a crazy manner) instead of a narrow band!
Time (t)
t = 0
t = 1
t = 2
t = 3
t = T

Growing Confidence

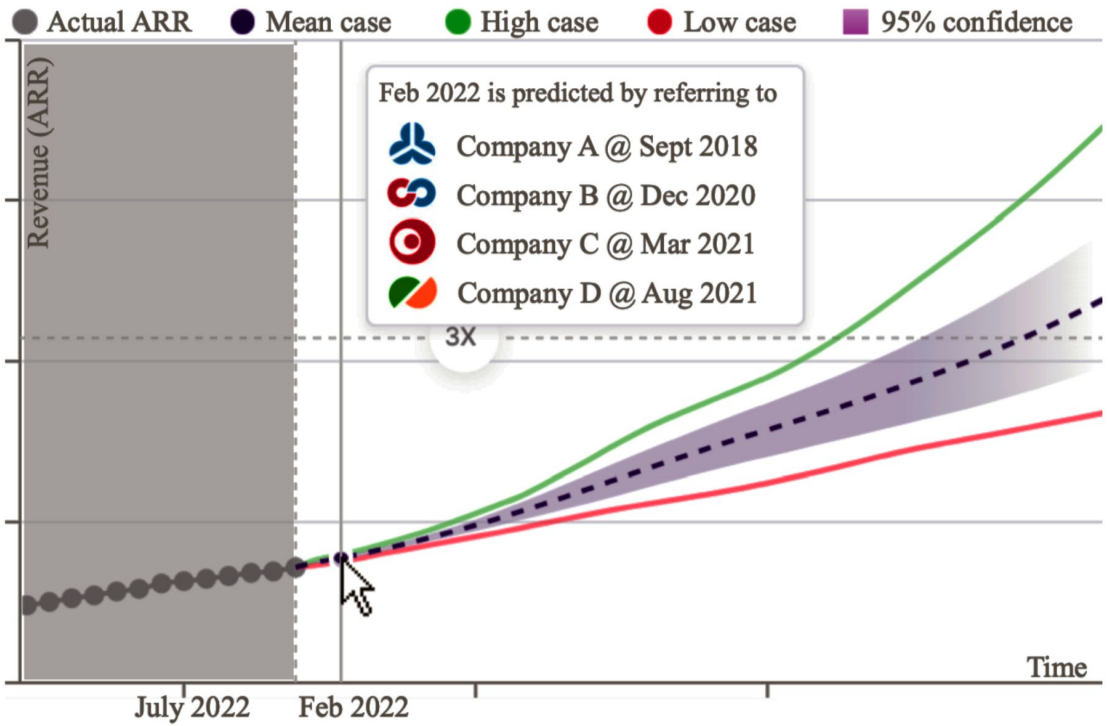


Extrapolate multiple times

&

Assume Gaussian

In Production!



Adapted from

EQT Motherbrain

Platform

Benchmarking

Datasets: ARR129 and SapiQ

Baselines: ARIMA, Prophet, DeepAR, LSTM and Informer

Evaluation procedure: "rolling origin" with 10 trajectories per baseline

Metrics using mean prediction: RMSE, MAPE and PCC

Metrics using confidence estimate prediction: NLL and ACC

Metrics:		RMSE		MAPE		PCC		NLL		ACC	
Dataset:		ARR129	SapiQ	ARR129	SapiQ	ARR129	SapiQ	ARR129	SapiQ	ARR129	SapiQ
Methods	SiRE (Ours)	**9.6917**	**57.8620**	**0.0480**	**0.6571**	**0.8284**	**0.6049**	**7.0578**	**8.5866**	**0.7102**	**0.5539**
	ARIMA [1, 3]	31.1630	117.0928	0.2091	0.9603	0.5590	0.4388	10.1357	10.6687	0.5230	0.3305
	Prophet [29]	33.0980	119.3963	0.3899	1.0763	0.5095	0.3780	9.9370	11.0289	0.5233	0.3203
	DeepAR [22]†	13.3720	76.1662	0.1347	0.8909	0.6212	0.5091	9.3044	9.8906	0.6300	0.4095
	LSTM [12]†	26.9251	88.0435	0.1894	0.9504	0.5721	0.4544	11.0396*	10.4210*	0.4983*	0.3407*
	Informer [39]†	12.7482	84.2029	0.0958	0.8630	0.7448	0.5207	9.5108*	10.2366*	0.6238*	0.4009*

Benchmarking

Datasets: ARR129 and SapiQ

Baselines: ARIMA, Prophet, DeepAR, LSTM and Informer

Evaluation procedure: "rolling origin" with 10 trajectories per baseline

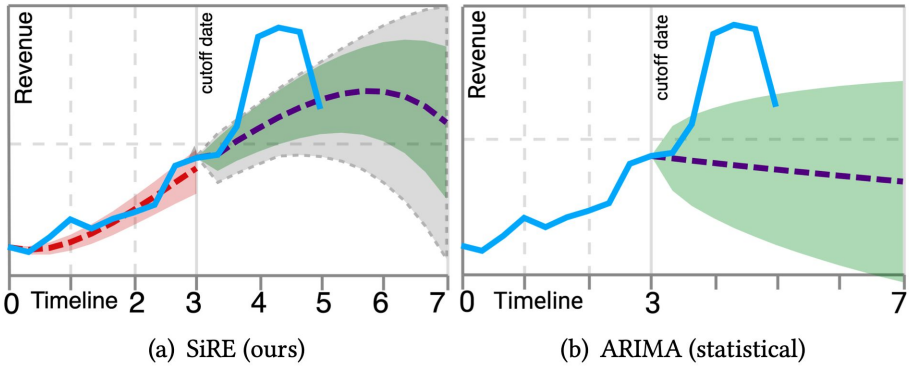
Metrics using mean prediction: RMSE, MAPE and PCC

Metrics using confidence estimate prediction: NLL and ACC

Metrics:		RMSE		MAPE		PCC		NLL		ACC	
Dataset:		ARR129	SapiQ	ARR129	SapiQ	ARR129	SapiQ	ARR129	SapiQ	ARR129	SapiQ
Methods	SiRE (Ours)	**9.6917**	**57.8620**	**0.0480**	**0.6571**	**0.8284**	**0.6049**	**7.0578**	**8.5866**	**0.7102**	**0.5539**
	ARIMA [1, 3]	31.1630	117.0928	0.2091	0.9603	0.5590	0.4388	10.1357	10.6687	0.5230	0.3305
	Prophet [29]	33.0980	119.3963	0.3899	1.0763	0.5095	0.3780	9.9370	11.0289	0.5233	0.3203
	DeepAR [22]†	13.3720	76.1662	0.1347	0.8909	0.6212	0.5091	9.3044	9.8906	0.6300	0.4095
	LSTM [12]†	26.9251	88.0435	0.1894	0.9504	0.5721	0.4544	11.0396*	10.4210*	0.4983*	0.3407*
	Informer [39]†	12.7482	84.2029	0.0958	0.8630	0.7448	0.5207	9.5108*	10.2366*	0.6238*	0.4009*

Qualitatively

- Missing data points are imputed.
- The trajectory is smooth so that patterns and trends are easier to identify.
- Confidence can be naturally estimated everywhere.
- The 95% Confidence Interval starts much narrower.



(a) SiRE (ours) (b) ARIMA (statistical)

(c) DeepAR (RNN based) (d) Informer (Transformer based)

Note: The actual name and revenue (Y-axis) of the scaleup is regarded as sensitive information and therefore removed from the plots. The forecast starts from the "cutoff date".

Investors Perspective

For example, will a company be valued 3x in 2 years?

We measure metrics like **TPR>2x_in_2/3y,** which is the true positive rate of scaleups that reach 2x revenue after 2 or 3 years.

Methods:		SiRE	ARIMA	DeepAR	Informer
ARR129	TPR>2x_in_2/3y	**0.6938**	0.5560	0.5904	0.5875
	TPR>3x_in_2/3y	**0.6809**	0.5100	0.6060	0.5723
SapiQ	TPR>2x_in_2/3y	**0.5780**	0.3903	0.4316	0.4126
	TPR>3x_in_2/3y	**0.5402**	0.3729	0.4040	0.3890
	TPR>3x_in_4/5y	**0.5537**	0.3505	0.3830	0.4057
	TPR>4x_in_4/5y	**0.5290**	0.3511	0.3714	0.3965

Thanks!

- For further questions and details, feel free to check out our code and paper: https://github.com/EQTPartners/sire
- Or, email us: tech_motherbrain-research@eqtpartners.com
- Learn more about EQT Motherbrain at: https://eqtgroup.com/motherbrain & https://eqtventures.com/motherbrain