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 is a highly relevant metric to evaluate a scaleup company!

- **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.

But,

- 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.

So, we need

a data-driven method that performs revenue extrapolation on scarce data in an automated way:

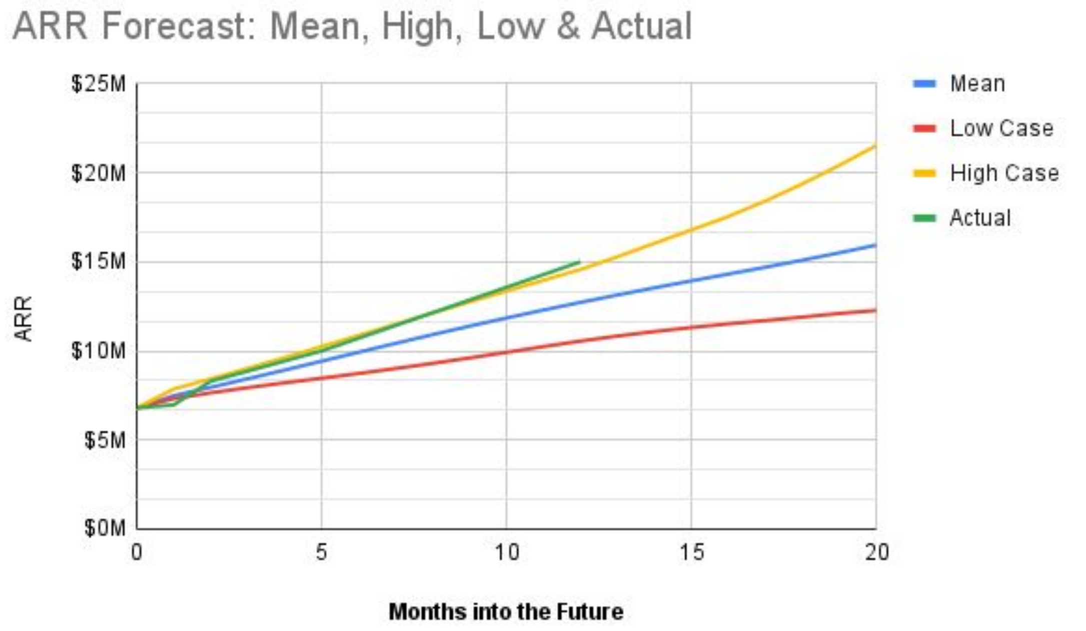
- A **quick way** to **assess companies' revenue potential** with little information needed;
- **Benchmarking** of a **manually produced revenue forecasting**.

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.

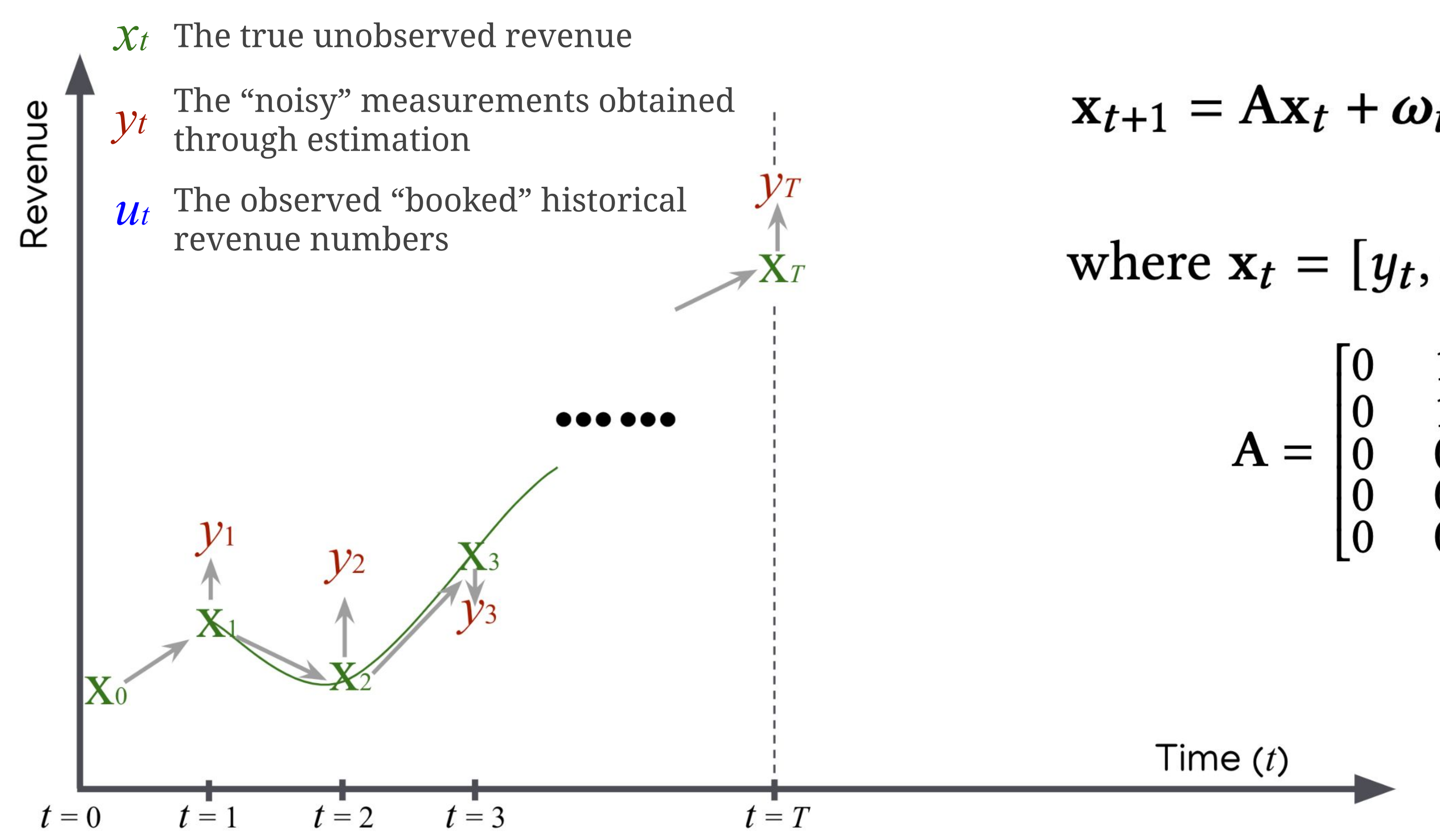
This is the first work that meets all practical requirements simultaneously.

An Example



The Revenue Model: A Linear Dynamical System

Model Parameter



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

$$\text{where } \mathbf{x}_t = [y_t, x_t, v_t, a_t, d_t]^\top \qquad \mathbf{c} = [1, 0, 0, 0, 0]$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1/2 & 1 \\ 0 & 1 & 1 & 1/2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad \omega_t \sim \mathcal{N}(0, \mathbf{Q}) \qquad \epsilon_t \sim \mathcal{N}(0, \mathbf{R})$$

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

Model Optimization: EM Algorithm

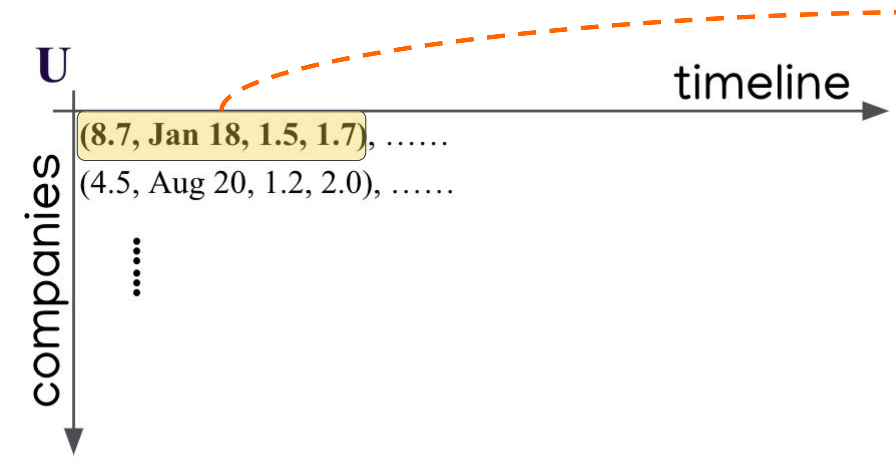
Target: locate x_t using observed measurement y_t

Measurement is the key!

Measurement: Measure y_{t+1} based on u_t using a small time-series dataset $\mathbf{U}$

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

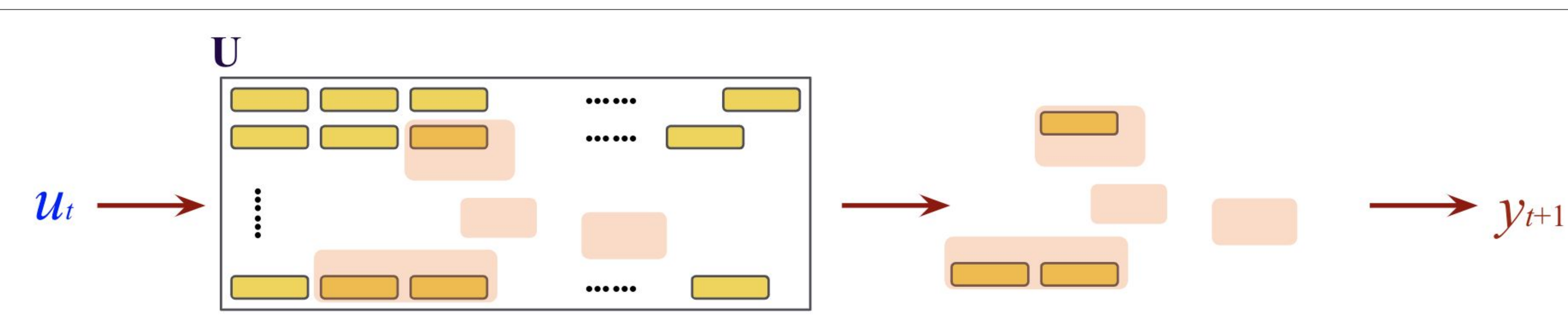
Dataset



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

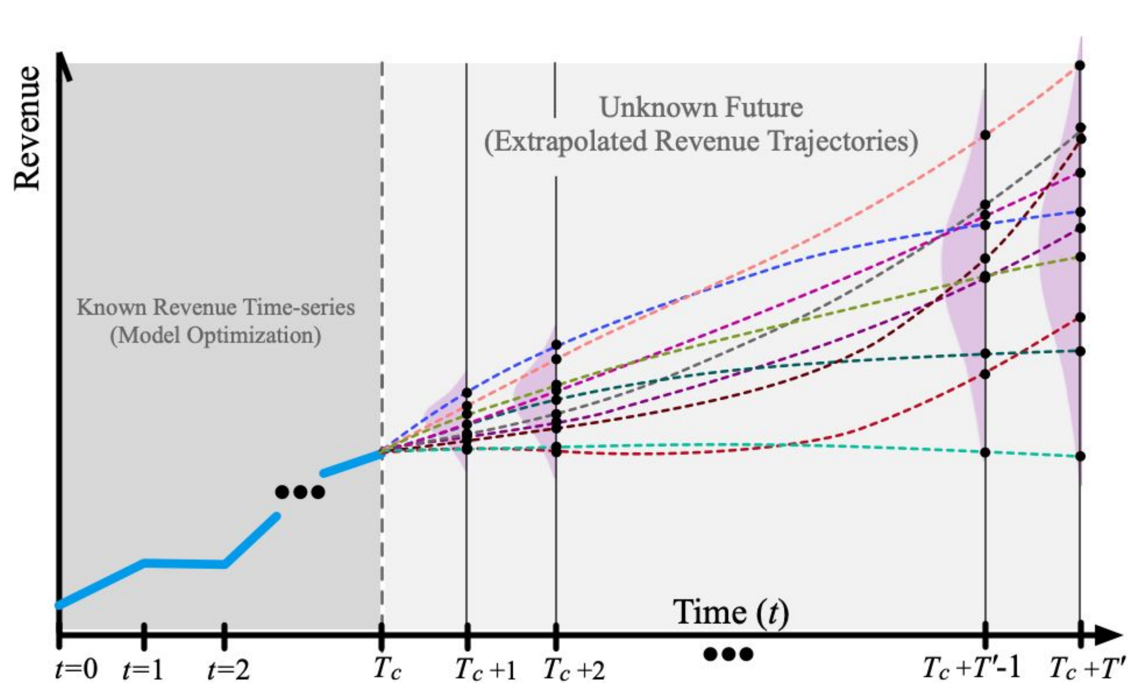
Procedure:



Filtering by:
business, year-month, revenue, YoY growth

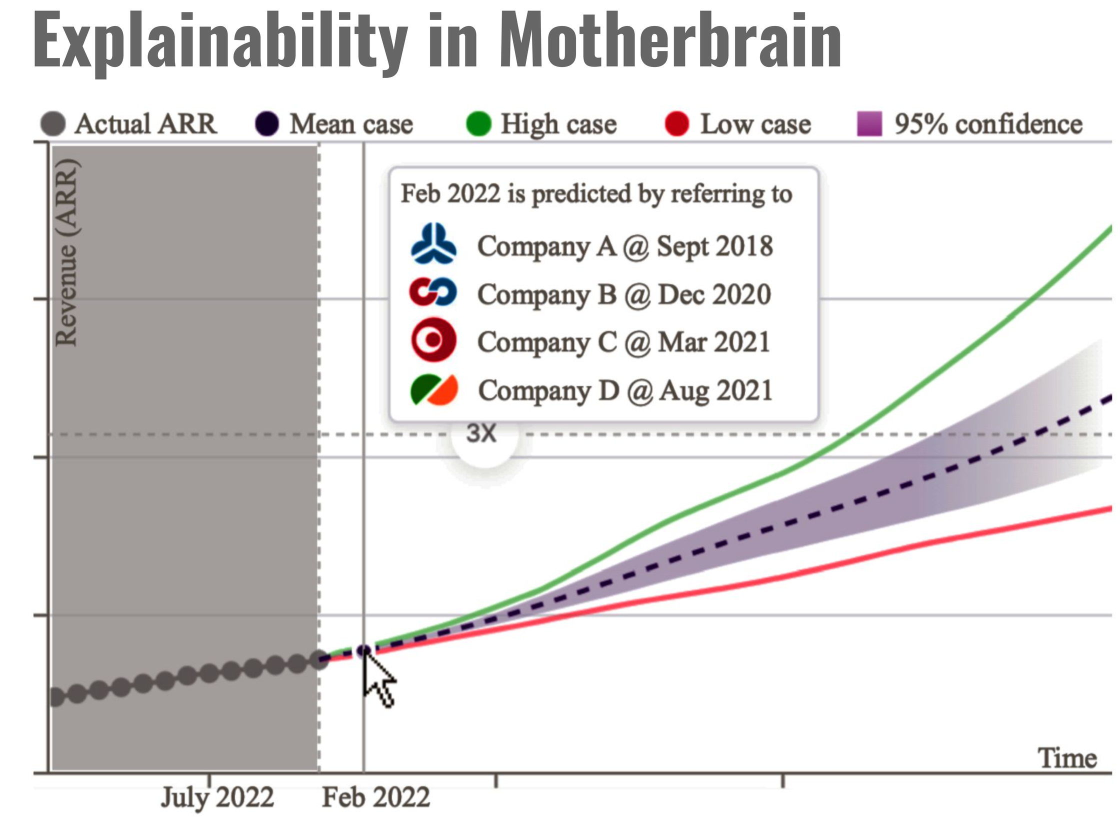
Sampling:
A **stochastic** approach!

Confidence Estimate



Extrapolate multiple times & Assume Gaussian

Explainability in Motherbrain



Benchmarking

2 datasets, 5 baselines, 5 evaluation metrics, "rolling origin" evaluation strategy

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*

Questions: lele.cao@eqtpartners.com

Source Code: https://github.com/EQTPartners/sire

Website: https://eqtgroup.com/motherbrain