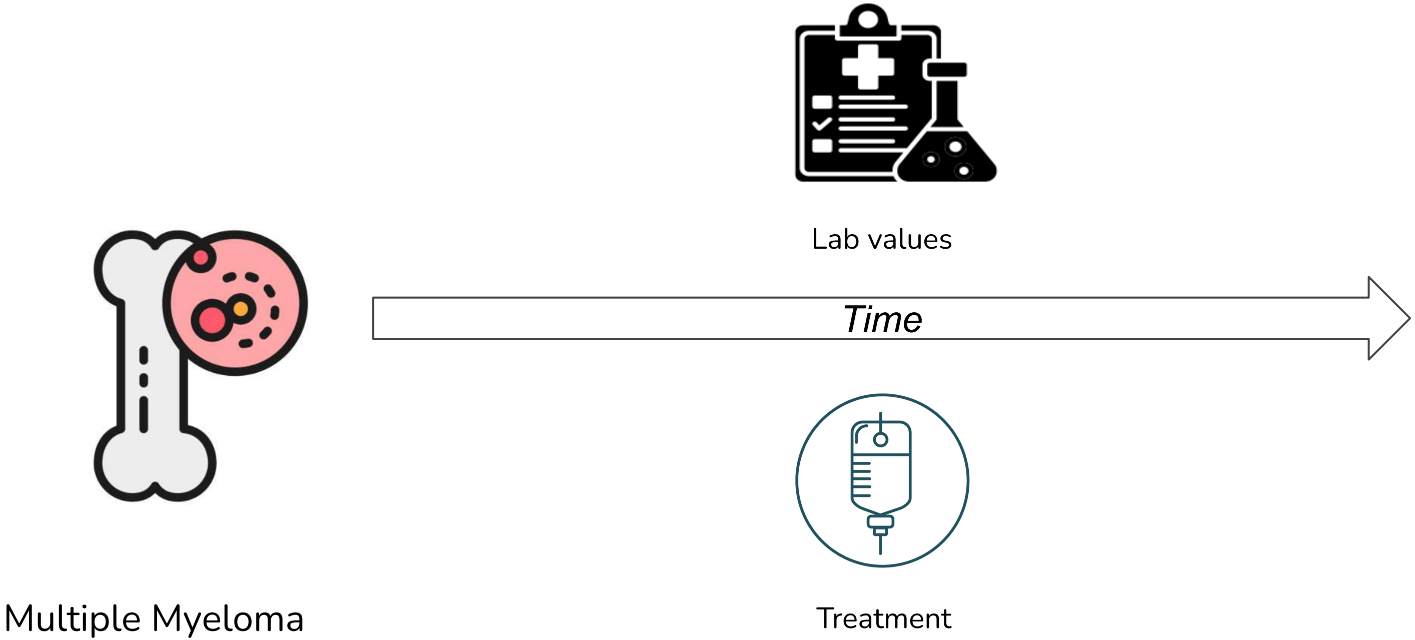
Time series analysis of biomarkers for multiple myeloma

Dianna Kan, Yang Qu, Qingyang Yu, Shujun Yan

Outline

- Motivation
- Research problem
- Data & preprocessing
- Univariate models
    - Background
    - Implementation
    - Result
- Multivariate models
- Next steps

Motivation



Lab values

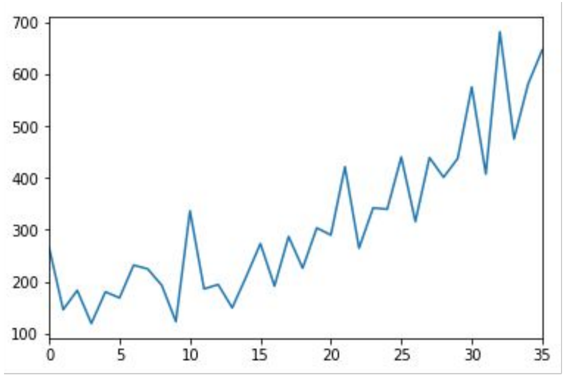
Time

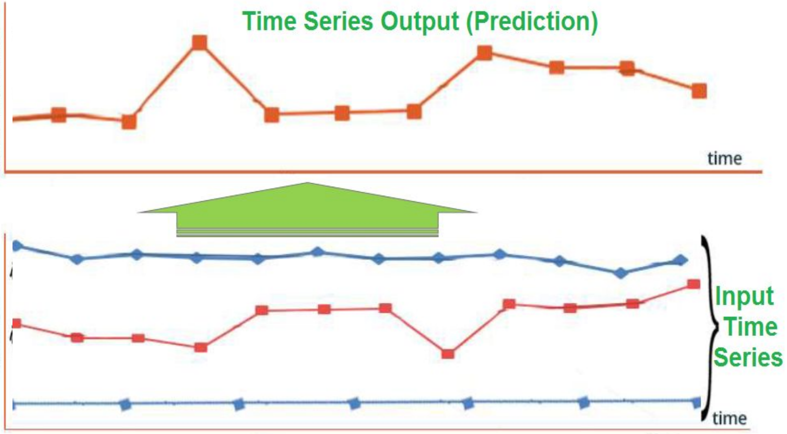
Multiple Myeloma

Treatment

Research Problem

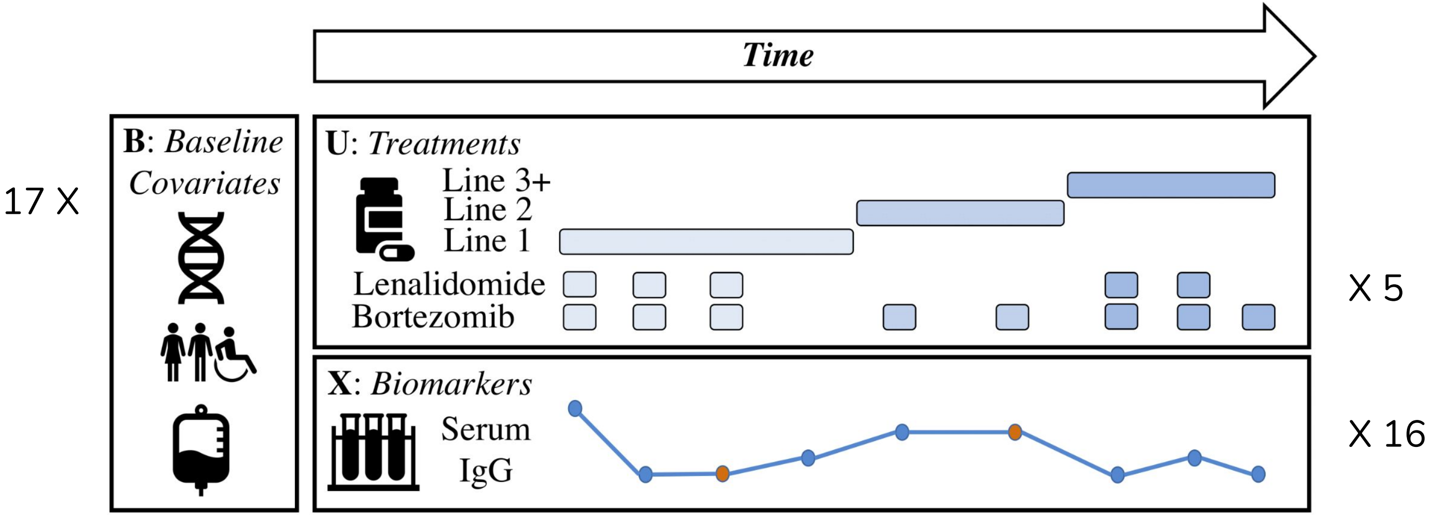
Univariate and multivariate time series analysis of lab values in the cohort of patients with multiple myeloma





Data & preprocessing

Machine Learning with Multiple Myeloma Research Foundation CoMMpass Dataset (ML-MMRF)



Machine Learning with Multiple Myeloma Research Foundation CoMMpass Dataset (ML-MMRF)
https://github.com/clinicalml/ml_mmrf

Univariate Models

Background - ARIMA

- Univariate model with 1D array-like time series input
- Autoregressive Integrated Moving Average(ARIMA)) model

$$Y_t = \alpha + \overset{\text{AR}}{\boxed{\phi_1 Y_{t-1} + ... + \phi_p Y_{t-p}}} + \overset{\text{MA}}{\boxed{\theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}}} + \epsilon_t$$

- Stationary Assumption: If not, use transformation like difference
- 3 key hyperparameters:
  - p: the number of lags of Y (Order of AR)
  - q: the number of lagged residuals (Order of MA)
  - d: the minimum number of difference for stationary
- Drawbacks:
  - Not consider effect of external variables
  - Only take one time series

Related works

- ARIMAX: ARIMA with exogenous variables

$$Y_t = \overset{\text{exogenous}}{\boxed{\gamma T_t}} + \overset{\text{ARIMA}}{\boxed{\alpha + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t}}$$

  - First point by Prof Rob J Hyndman on his website in 2010

- Application of ARIMAX

  - Thailand Export study: ARIMAX outperforms ARIMA

    Exogenous: trade partners’ Composite Leading Indicator (CLI)

- Still just one time series analysis

https://www.researchgate.net/publication/255731345_Autoregressive_Integrated_Moving_Average_with_Explanatory_Variable_ARIMAX_Model_for_Thailand_Export

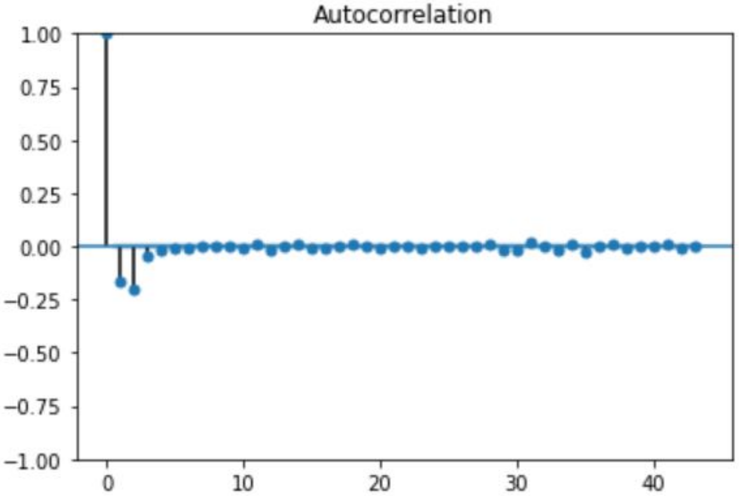
ARMA

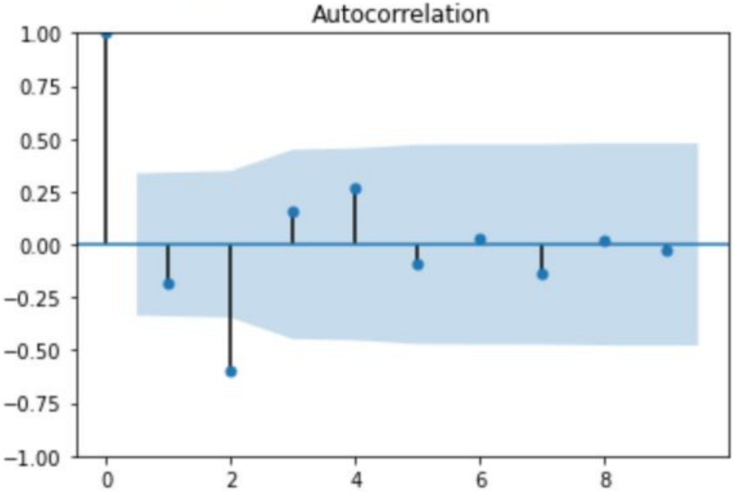
Augmented Dickey Fuller test

d=0

All patients: $p=0.0 < 0.05$

1 patient: $p=0.016 <0.05$

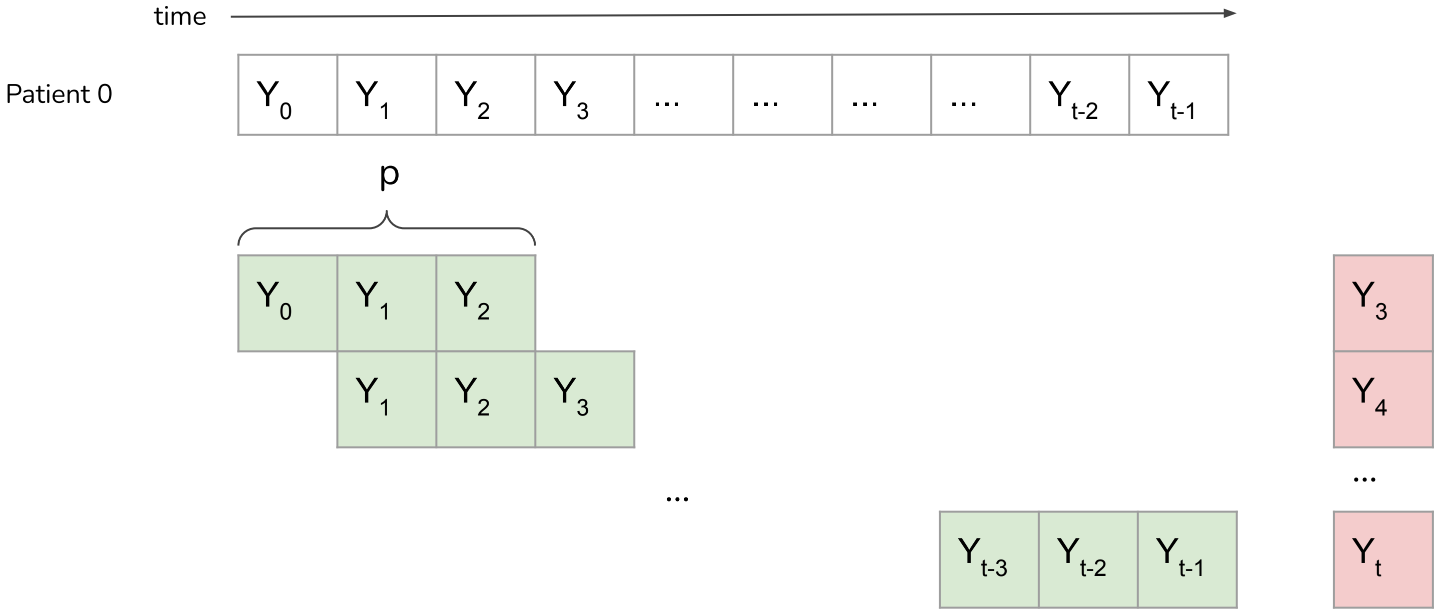




ARMA - AR

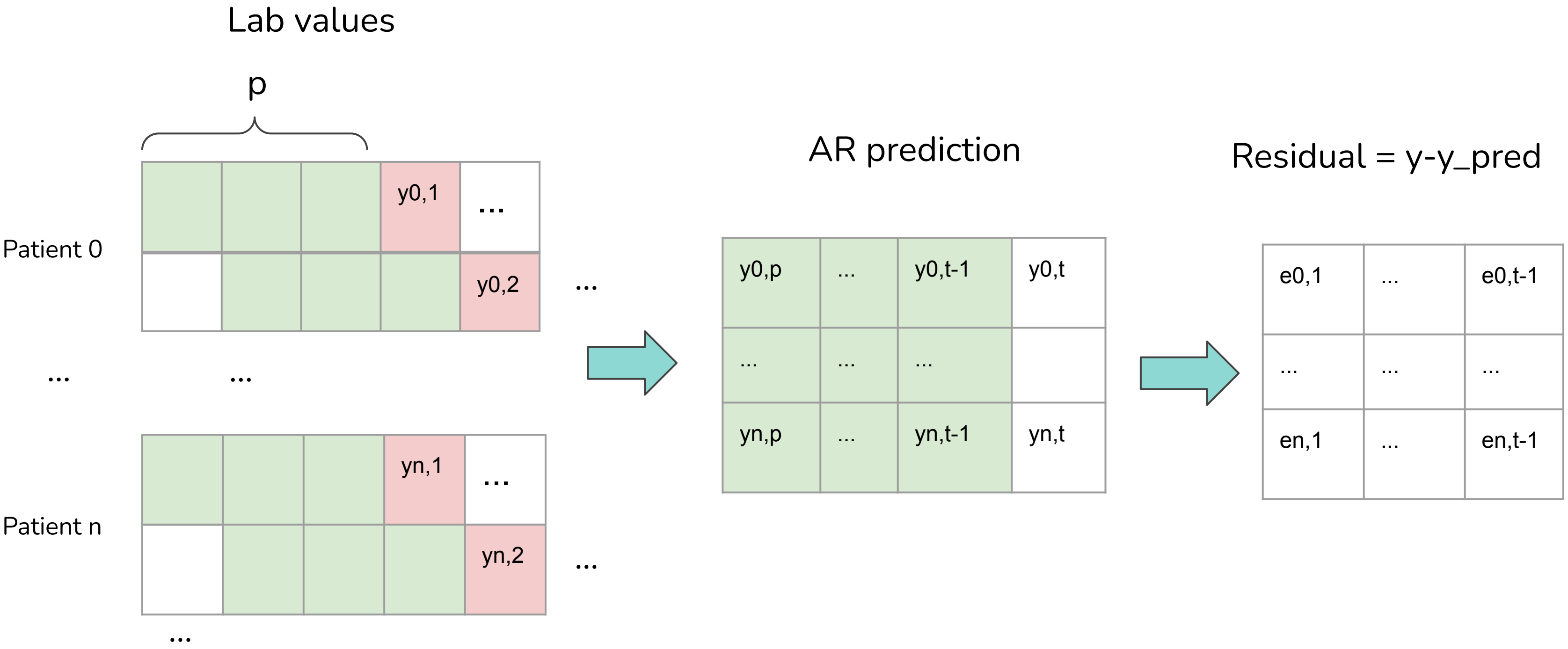
$$Y_t = \boxed{\alpha + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p}} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} + \epsilon_t$$

Lab values


time
Patient 0
Y_0 Y_1 Y_2 Y_3 ... ... Y_{t-2} Y_{t-1}
p
Y_0 Y_1 Y_2
Y_1 Y_2 Y_3
...
Y_{t-3} Y_{t-2} Y_{t-1}
Y_3
Y_4
...
Y_t

ARMA - AR

$$Y_t = \boxed{\alpha + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p}} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} + \epsilon_t$$



🟩	🟩	🟩	y0,1	...
	🟩	🟩	🟩	y0,2

🟩	🟩	🟩	yn,1	...
	🟩	🟩	🟩	yn,2

AR prediction

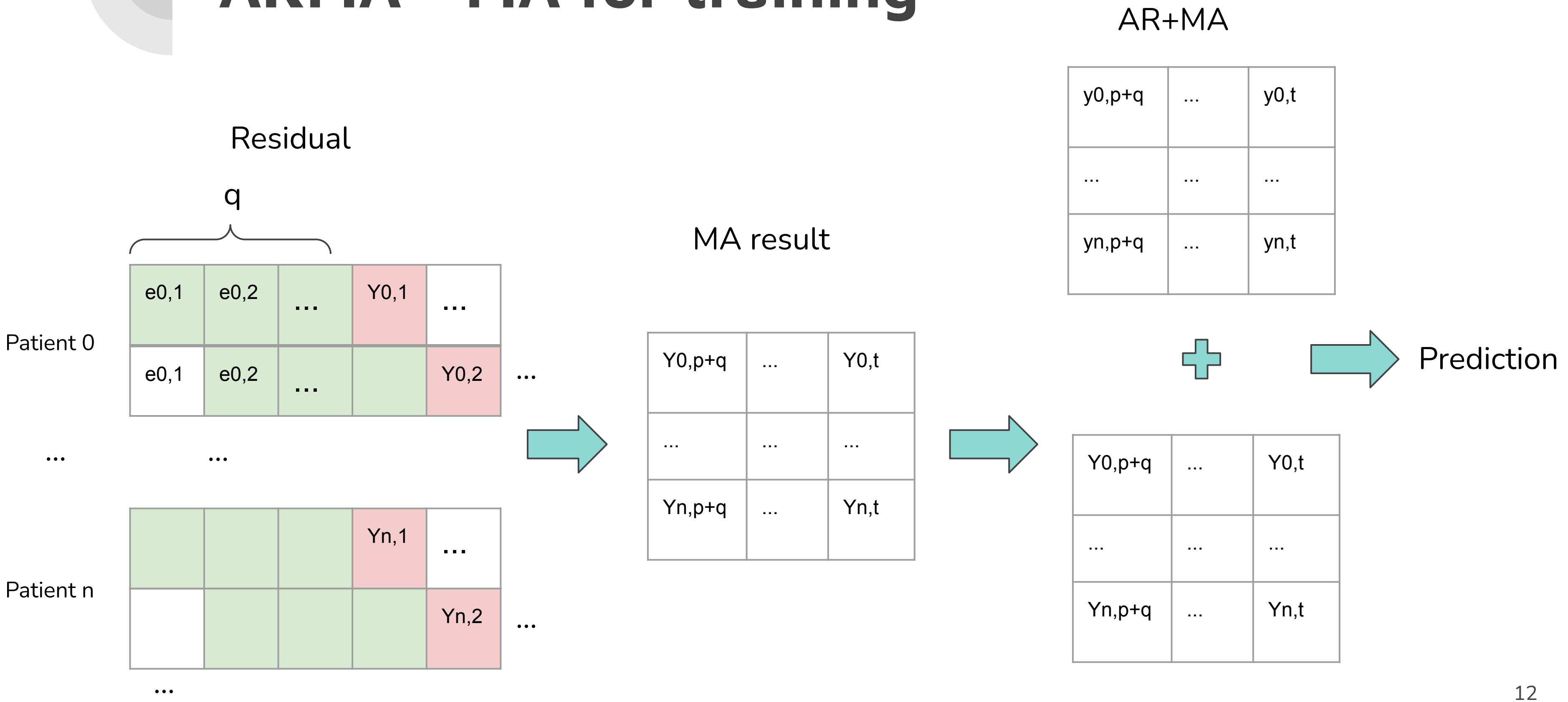
y0,p	...	y0,t-1	y0,t
...	...	...	
yn,p	...	yn,t-1	yn,t

Residual = y-y_pred

e0,1	...	e0,t-1
...	...	...
en,1	...	en,t-1

$$Y_t = \alpha + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \boxed{\theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} + \epsilon_t}$$

ARMA - MA for training



Residual

q

Patient 0	e0,1	e0,2	...	Y0,1	...	
	e0,1	e0,2	...	🟩	Y0,2	...
...	...					
Patient n	🟩	🟩	🟩	Yn,1	...	
		🟩	🟩	🟩	Yn,2	...
	...					

MA result

Y0,p+q	...	Y0,t
...	...	...
Yn,p+q	...	Yn,t

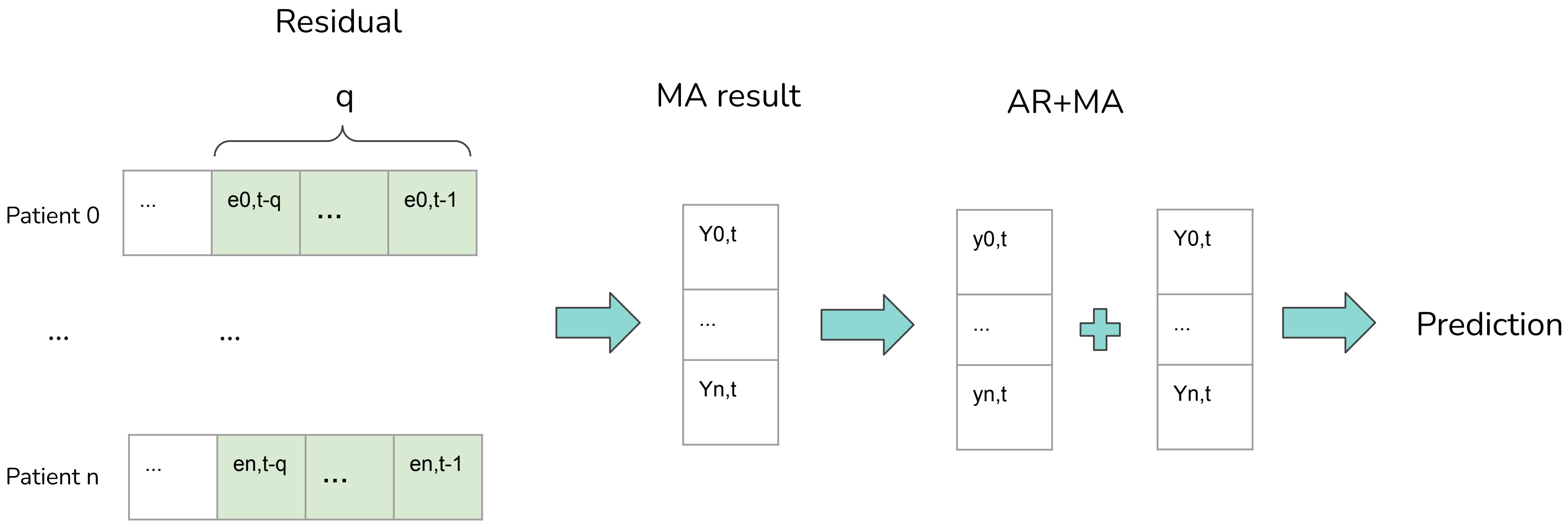
AR+MA

y0,p+q	...	y0,t
...	...	...
yn,p+q	...	yn,t

Y0,p+q	...	Y0,t
...	...	...
Yn,p+q	...	Yn,t

Prediction

ARMA - MA for Prediction



Residual

q

Patient 0

...	e0,t-q	...	e0,t-1

...

...

Patient n

...	en,t-q	...	en,t-1

MA result

Y0,t

...

Yn,t

AR+MA

y0,t

...

yn,t

Y0,t

...

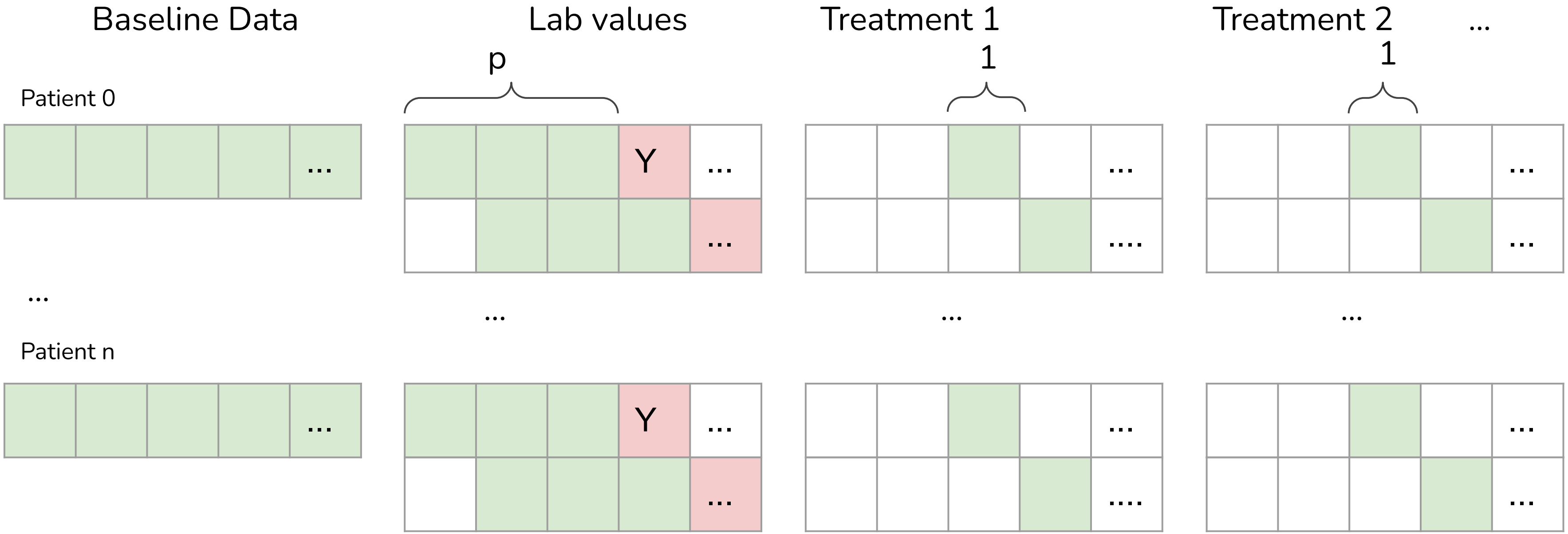
Yn,t

Prediction

BARMA

$$Y_t = \alpha + \overset{\text{AR}}{\boxed{\phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p}}} + \overset{\text{Baseline}}{\boxed{\beta X}} + \overset{\text{Treatment}}{\boxed{\gamma T_{t-1}}} + \overset{\text{MA}}{\boxed{\theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}}} + \epsilon_t$$

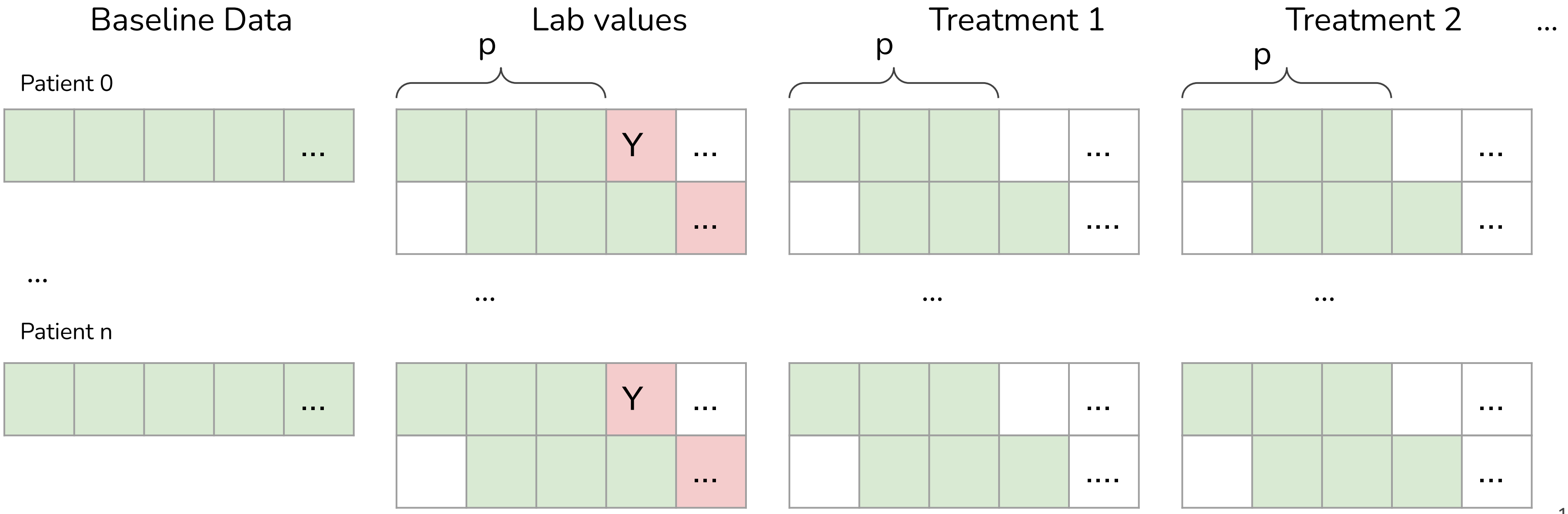
- ARMA + Baseline + Most recent treatment at time t-1 (just one timestamp)



BARMAX

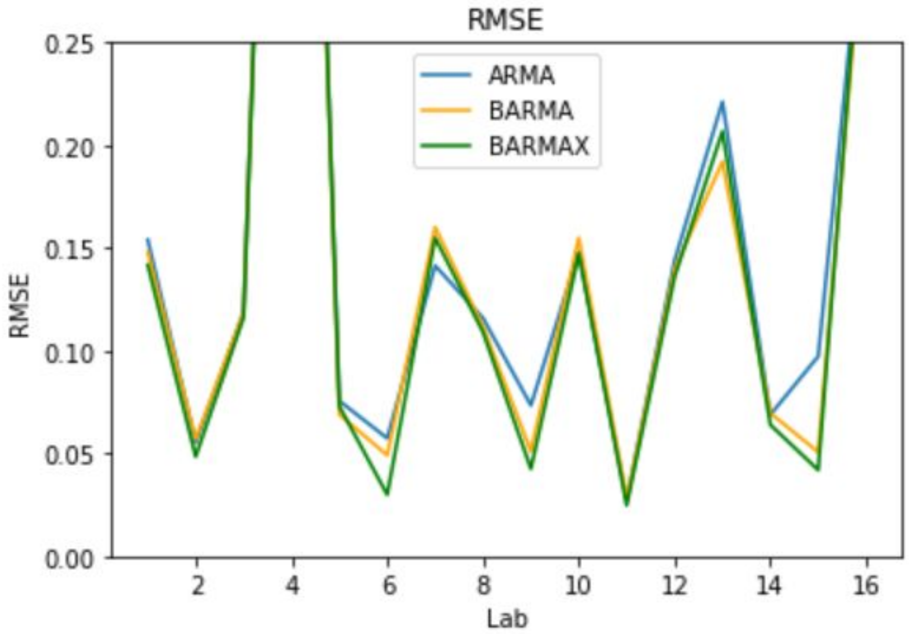
- ARMA + Baseline + AR Treatment

$$Y_t = \alpha + \overset{\text{AR}}{\boxed{\phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p}}} + \overset{\text{Baseline}}{\boxed{\beta X}} + \overset{\text{Treatment}}{\boxed{\gamma_1 T_{t-1} + \ldots + \gamma_p T_{t-p}}} + \overset{\text{MA}}{\boxed{\theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t}}$$



Results

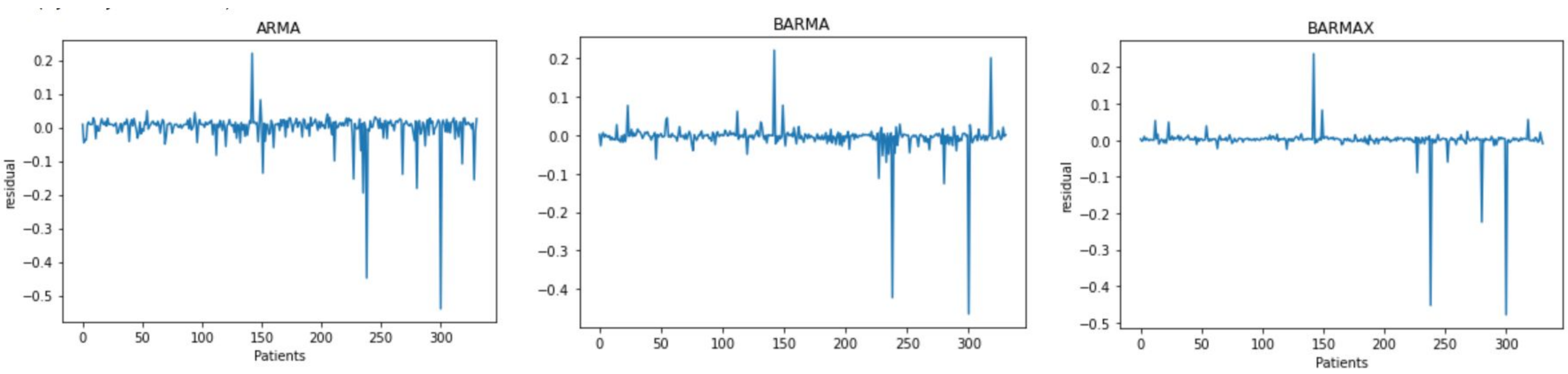
Model	Number of features
ARMA	p + q
BARMA	p + q + 8 + 62
BARMAX	p + q + p*8 + 62



Tuning p,q:
5-fold cross validation

Results

Prediction Residual of Lab 15 serum_igm across all patients



RMSE: 0.09728

RMSE: 0.05065

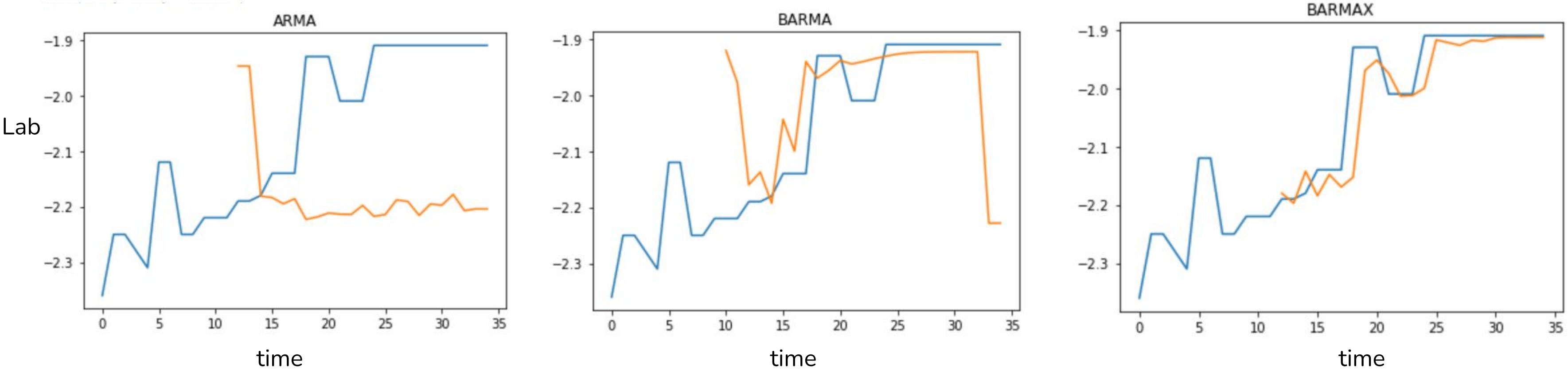
RMSE: 0.04193

Results

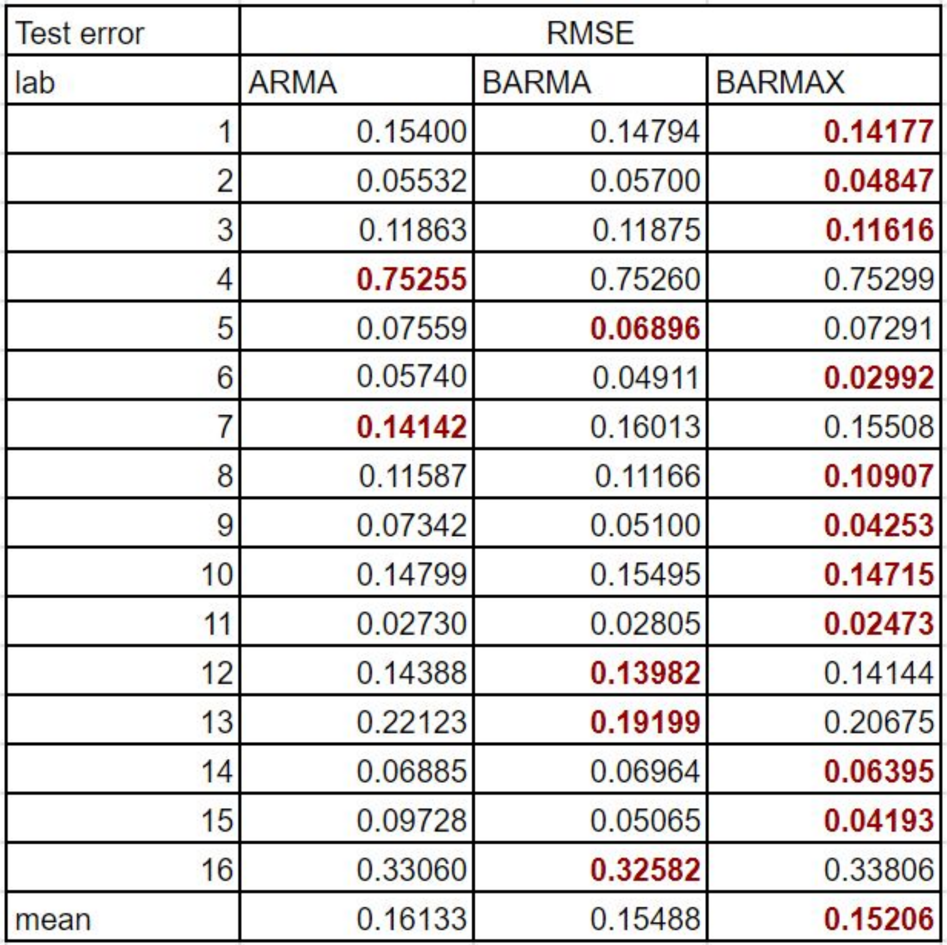
— true

— predicted

Lab 15 serum_igm Prediction of 50th Patient



Results

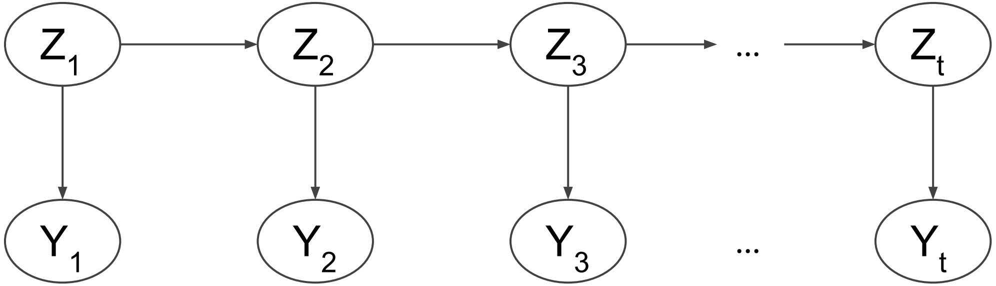


Test error	RMSE		
lab	ARMA	BARMA	BARMAX
1	0.15400	0.14794	**0.14177**
2	0.05532	0.05700	**0.04847**
3	0.11863	0.11875	**0.11616**
4	**0.75255**	0.75260	0.75299
5	0.07559	**0.06896**	0.07291
6	0.05740	0.04911	**0.02992**
7	**0.14142**	0.16013	0.15508
8	0.11587	0.11166	**0.10907**
9	0.07342	0.05100	**0.04253**
10	0.14799	0.15495	**0.14715**
11	0.02730	0.02805	**0.02473**
12	0.14388	**0.13982**	0.14144
13	0.22123	**0.19199**	0.20675
14	0.06885	0.06964	**0.06395**
15	0.09728	0.05065	**0.04193**
16	0.33060	**0.32582**	0.33806
mean	0.16133	0.15488	**0.15206**

Multivariate Models

Hidden Markov Model- Baseline

Z_i: hidden state at time t
Y_i: observed lab values at time t



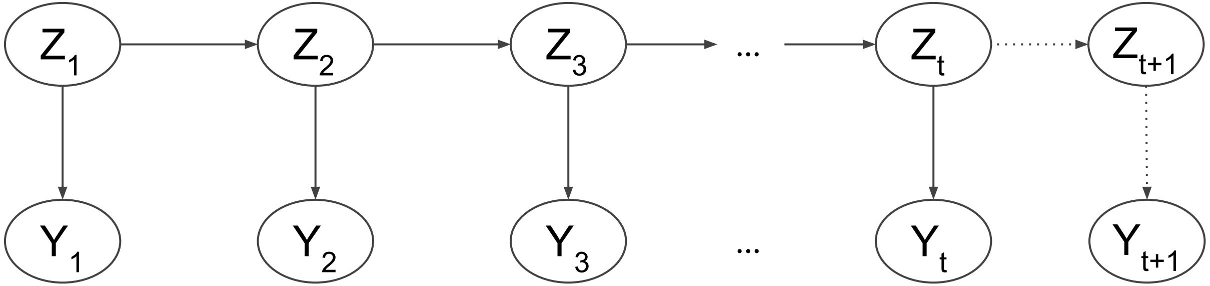
Transition probability a_{ij}: the probability of moving from state i to state j underlying the Markov chain

Emission probability b_i: the probability of observing the lab values given the hidden state at time i

Hidden Markov Model - Baseline

Z_i: hidden state at time i

Y_i: observed lab values at time t

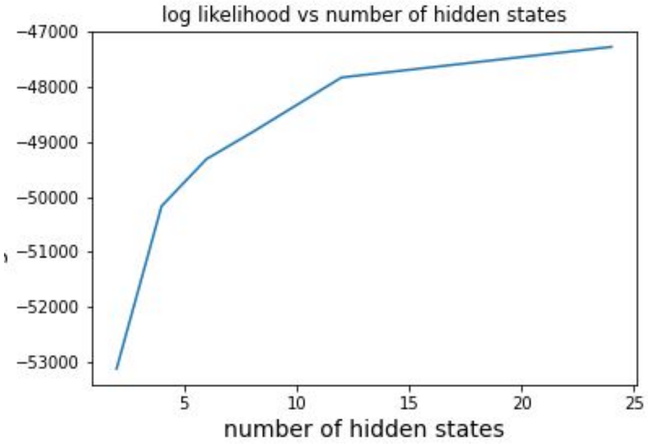


Prediction:

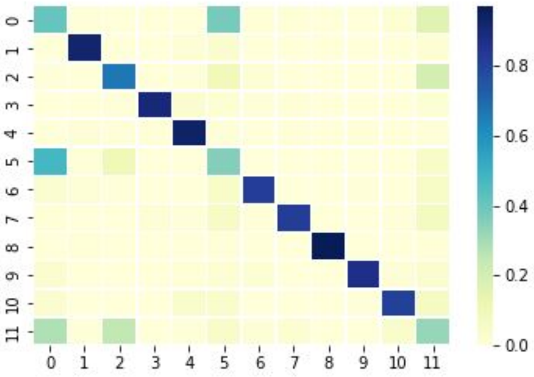
- Predict the hidden state at the next time point
- Compute the mean and variance of each feature at each hidden state
- Draw a sample from Gaussian distribution to be the predicted value

Hidden Markov Model - Baseline

- Number of hidden state
  - Tuned based on log-likelihood
  - Choose 12 hidden states



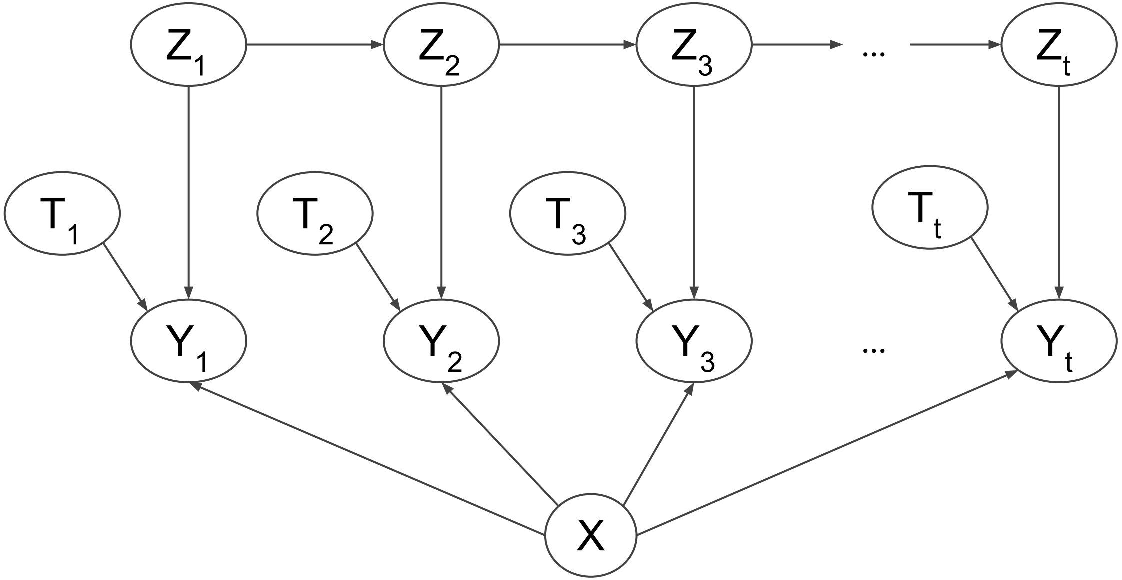
- Transition matrix



Hidden Markov Model - Advanced

- Hidden Markov Model - Baseline
  - only takes observed lab values into account
- But we have additional information
  - Treatment data
  - Baseline covariates
- Hidden Markov Model - Advanced
  - Modified the baseline model to incorporate with the above information
  - Assume:
    - Treatment and baseline covariates are independent of hidden state
    - Treatment will only affect the next observed lab value

Hidden Markov Model - Advanced



Z_i: hidden state at time i
Y_i: observed lab values at time t
T_i: treatment at time i
X: baseline covariate

Next steps

- Continue to work on implementation of Hidden Markov Model with treatments and baseline covariates
- Compare results between:
    - Baseline Hidden Markov Model
    - Hidden Markov Model with treatments and baseline covariates
- Compare results between univariate vs. multivariate model

Conclusion

Strength:

- Modified ARIMA & Hidden Markov Model to incorporate external variables (baseline covariates, treatment data)
- BARMAX: Good interpretability

Limitations:

- Missing values in the dataset - standardization/imputation issue
- Assumption on treatment effect on the lab values at the next time point
- Assumptions in modified Hidden Markov Model

Future Work

BARMAX:

- Independent lag order for treatment
- Regularization
- Toggle treatments to test potential future lab value outcomes

Hidden Markov Model:

- Deep Markov Model

Thank you!