Mini-Project (ML for Time Series) - MVA 2024/2025

Soël Megdoud soel.megdoud@ens-paris-saclay.fr
Julien Delavande julien.delavande@ens-paris-saclay.fr

May 22, 2025

1 Introduction and Contributions

In this mini-project, we focus on learning a meaningful graph representation from data, following the framework outlined by Dong *et al.* [1], and then leveraging this learned graph for time-series forecasting. The original article discusses different strategies for inferring graph topologies from observed signals, highlighting a *graph signal processing* (GSP) perspective.

Our Data and Objectives. To test these models, we use **COVID-19 weekly case counts** from multiple countries. Our goal is twofold:

- Learn a graph that best represents the relationships among different countries.
- Evaluate how graph-based models (AR+diffusion and SVAR) can improve time-series forecasting accuracy over simpler baselines.

Team Organization.

- *Julien*: Implementation and experiments with the AR+diffusion model, reading and summarizing the article [1].
- *Soël*: Implementation of the SVAR model (adjacency matrix, polynomial expansions), extensive hyperparameter tuning, and final result analysis.

Code Reuse. We reused standard optimization tools (`scipy.optimize.minimize`), matrix exponentials (`scipy.linalg.expm`), and typical data scaling classes (`MinMaxScaler`, `StandardScaler`). About 30% of our code is adapted from open-source scripts; 70% was developed specifically for this project.

Extension Beyond the Article. We propose an additional **AR+diffusion** model, not explicitly mentioned in the original article, to forecast time series on the nodes of a learned graph. We further compare this approach with a **Structural Vector Autoregression (SVAR)** model, where polynomial expansions of an adjacency matrix capture temporal dynamics.

2 Method

Our methodology combines **graph learning** with two distinct forecasting models, *AR+diffusion* and *SVAR*, both leveraging the inferred graph structure.

2.1 AR+Diffusion Model

We introduce an **AR(p)** [2] term combined with a **diffusion** term over the graph using the idea from [3] but using the previous time step as the source of the diffusion kernel and adding an AR model. The main idea is that the AR model is better at modeling the evolution of cases within a country, whereas the diffusion model is better at modeling external effects linked to other countries:

$$X_i(t) = \sum_{k=1}^{p} \alpha_k \, X_i(t-k) \; + \; \beta \left[\exp(-\tau \mathbf{L}) \, \mathbf{X}(t-1) \right]_i,$$
$$\mathbf{X}(t) \in \mathbb{R}^N, \quad i = 1, \ldots, N,$$

where $\exp(-\tau\mathbf{L})$ models how each node influences others through the learned graph $\mathbf{L}$, a Laplacian matrix. The parameters α_k control the autoregressive contributions, β the diffusion strength, and τ the intensity of the diffusion process. This model is flexible, allowing pure autoregressive behavior when $\tau = 0$ or pure diffusion with $p = 0$.

The training process minimizes a loss function defined as:

$$\text{Loss} = \sqrt{\sum_{t=p}^{K} \|\mathbf{X}(t) - \hat{\mathbf{X}}(t)\|^2} + \lambda \|\mathbf{L}\|, \tag{1}$$

where λ is the regularization coefficient, and $\|\mathbf{L}\|$ can represent either ℓ_1 (sparsity) or ℓ_2 (smoothness) norms on the off-diagonal elements of $\mathbf{L}$.

The optimization is performed using gradient-based methods such as `L-BFGS-B` from `scipy.optimize`.

2.2 SVAR Model

The **Structural Vector Autoregression (SVAR)** model [4], [5] which incorporates a polynomial expansion of the adjacency matrix $\mathbf{A}$ to capture temporal dynamics [6]. For a lag order M, the model predicts $\mathbf{x}_t$ as:

$$\mathbf{x}_t \approx \sum_{m=1}^{M} \left(\sum_{j=0}^{m} c_{m,j} \, \mathbf{A}^j \right) \mathbf{x}_{t-m}, \tag{2}$$

where $c_{m,j}$ are polynomial coefficients. These coefficients, along with $\mathbf{A}$, are learned by minimizing the loss:

$$\text{Loss} = \frac{1}{2} \sum_{t=M+1}^{K} \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|^2 + \text{Reg}_{\mathbf{A}}(\lambda_1) + \text{Reg}_{\mathbf{c}}(\lambda_2), \tag{3}$$

with ℓ_1 or ℓ_2 regularization applied to $\mathbf{A}$ and $\mathbf{c}$, λ_1 and λ_2 being the regularization coefficients on A and c respectively.

The residual is computed by iteratively subtracting polynomial expansions from the observed series $\mathbf{x}_t$, implementing the simplified algorithm described in [6]. Optimization is performed using `scipy.optimize.minimize` with methods like `L-BFGS-B` or `TNC` for efficiency.

Both models were implemented to handle various graph structures and regularization settings, providing insights into their trade-offs and practical applications. Forecasting performance (so called error in the following) was evaluated using the L2 norm between the predicted and true vectors.

3 Data

COVID-19 Weekly Data. We downloaded country-level COVID-19 case counts from Our World in Data (`https://github.com/owid/covid-19-data`). We selected 22 countries from different continents to ensure variety in epidemiological patterns. We then:

- **Aggregated daily data to weekly** to smooth out short-term oscillations.
- Imputed missing data with zeros.

Scaling and Splits. We used the `MinMaxScaler` to fit on the first 80% of the time steps (training set). The last 20% of weeks formed the test set. Countries with large disparities in COVID-19 scales, such as the United States vs. smaller nations, motivated a robust scaling approach.

Data Diagnostics. The data exhibits a clear weekly periodicity across different countries, likely due to variations in workdays and reporting practices. To mitigate this effect, we applied a smoothing technique to aggregate data over weekly intervals. Additionally, further diagnostics could involve identifying outliers or inconsistencies in reporting patterns, as these could bias the results. 3

From a modeling perspective, we assume that the dynamics are primarily driven by the temporal evolution of cases within and between countries. However, these assumptions may overlook structural biases in the data, such as underreporting or differences in testing rates. For instance, countries with similar cultural or economic characteristics may conduct testing campaigns simultaneously, creating artificial correlations. Future data analysis could investigate these patterns more deeply, incorporating supplementary datasets like mobility data or healthcare infrastructure to validate or challenge these assumptions.

4 Results

4.1 AR+Diffusion Forecasting

We tested $p \in \{1, \ldots, 5\}$, $\tau \in \{0, 0.01, 0.1, 0.5, 1, 2, 5, 10\}$, and regularization levels $\lambda_{\text{reg}} \in \{0, 0.01, 0.1, 1, 10\}$.

The first conclusion is that the pure diffusion model ($p = 0$, error$= 0.28$) performs worse than the pure autoregressive model ($\tau = 0$, error$= 0.24$), highlighting that within-country dependencies outweigh inter-country interactions. This observation motivated the inclusion of the AR term in the combined model. However, even the best pure AR model ($p = 3$) was outperformed by the combined AR+Diffusion model with parameters $p = 3$ and $\tau = 10^{-3}$, (error$= 0.23$). 4

Interestingly, the parameter τ does not significantly impact performance except when it is too large (e.g., $\tau = 10$), where the results degrade. This suggests that moderate levels of diffusion effectively complement autoregressive dynamics without overwhelming them.

Finally, examining the learned graph reveals that the strongest connections are not necessarily tied to geographical proximity. For instance, cultural dependencies were observed between England, Australia, and the United States. Conversely, Japan and Corea appeared relatively isolated, which may reflect their strict border control policies during the COVID-19 pandemic. 1

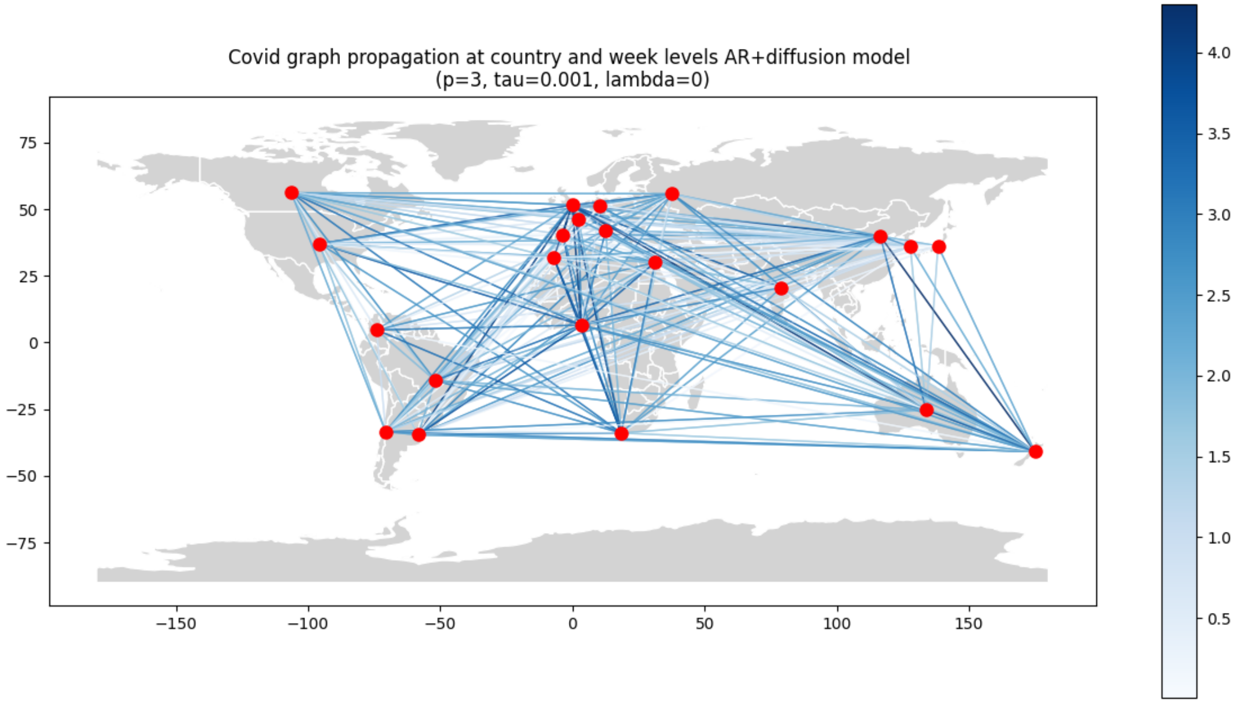

Figure 1: AR + Diffusion model - graph

4.2 SVAR Forecasting

Similarly, for the SVAR approach, we performed a grid search over $M \in \{1, \ldots, 5\}$ and regularization coefficients $\lambda_1, \lambda_2 \in \{0, 0.01, 0.1, 1, 10\}$ for either L1 or L2.

The results indicate that the impact of M (lag order) stabilizes after $M = 2$, with a slight underperformance at $M = 1$. From $M = 2$ onward, the SVAR model consistently outperformed the $t - 1$ baseline (error$=$ 0.26), pure diffusion, pure AR, and the AR+Diffusion model discussed earlier. For $M = 4$, SVAR emerged as the best-performing model on this dataset (error$=$ 0.22). 5

The learned graph is notably sparse, with stronger connections observed between specific countries. Interestingly, these connections are not always intuitive or geographically driven. For example, a strong link between South Africa and Colombia may be explained by economic similarities, as both are emerging markets with potentially comparable public health measures during the pandemic. Additionally, strong links were observed between these countries and others in Africa, such as Egypt, Morocco, and Nigeria, highlighting shared dynamics within the region. This sparsity and focus on key relationships underscore the SVAR model's ability to capture meaningful temporal and structural patterns. 2

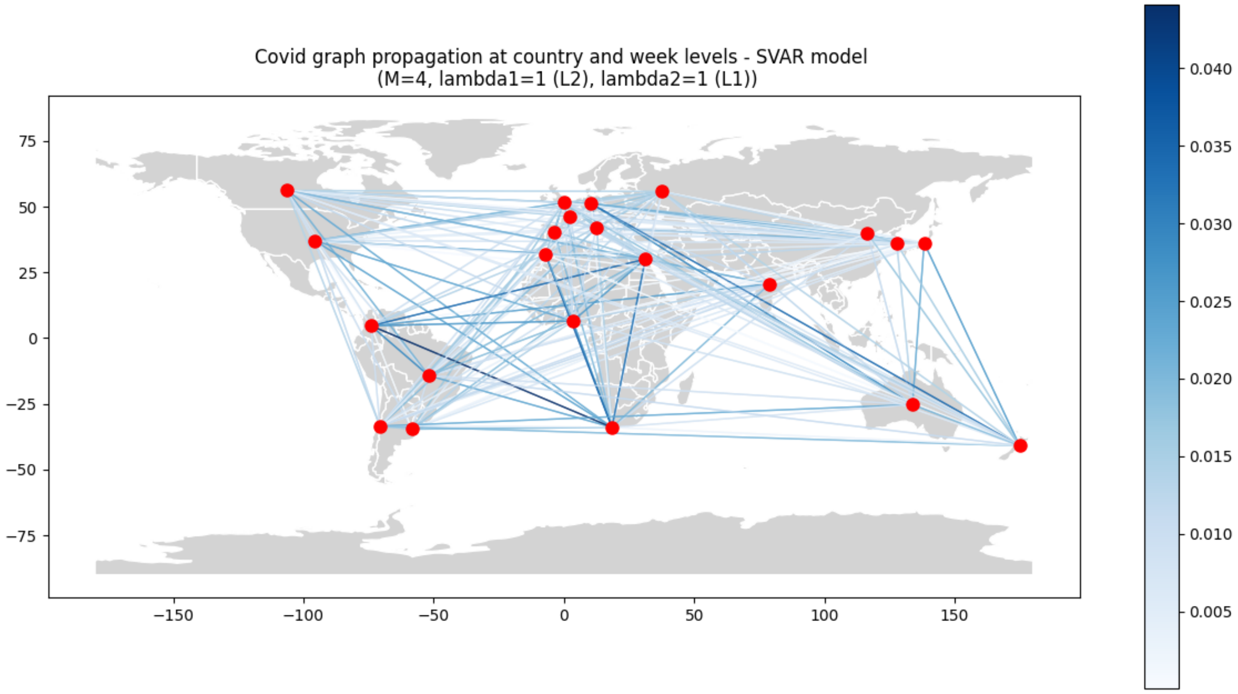

Figure 2: SVAR model - graph

4.3 Concluding Remarks

The best-performing model across all experiments was the SVAR model with $M = 4$, which consistently outperformed other approaches, including pure AR, pure diffusion, and the combined AR+Diffusion model. However, interpreting the learned graphs remains challenging, as the connections are not always geographically intuitive. This may reflect factors such as cultural or economic similarities, or synchronized public health measures, which lead to similar patterns in the time series data.

There are notable limitations to our approach. First, the nodes in our models represent countries, which may not fully capture the dynamics of epidemic diffusion at the individual level. A more granular representation with individuals as nodes would provide a better understanding of within-population diffusion processes. Second, the reliability of reported COVID-19 case data varies across countries, potentially introducing biases. Factors such as underreporting, differences in testing strategies, and synchronous testing campaigns in culturally aligned countries can influence the observed data.

For the SVAR model specifically, the reliance on a single adjacency matrix may oversimplify the temporal evolution of relationships. Future work could explore dynamic graphs that evolve over time or incorporate additional contextual data (e.g., mobility or vaccination rates) to refine the model's predictions. Additionally, extending the methodology to individual-level data could provide a more nuanced understanding of epidemic spread and improve the applicability of these models in public health scenarios.

References

[1] Xiaowen Dong, Dorina Thanou, Michael Rabbat, and Pascal Frossard. Learning graphs from data: A signal representation perspective. *IEEE Signal Processing Magazine*, 36(3):44–63, 2019.

[2] G. E. P. Box and G. M. Jenkins, “Time Series Analysis: Forecasting and Control,” *Holden-Day*, 1st ed., 1970.

[3] D. Thanou, X. Dong, D. Kressner, and P. Frossard, “Learning heat diffusion graphs,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 484–499, 2017.

[4] J. Songsiri and L. Vandenberghe, “Topology selection in graphical models of autoregressive processes,” *Journal of Machine Learning Research*, vol. 11, pp. 2671–2705, 2010.

[5] A. Bolstad, B. D. V. Veen, and R. Nowak, “Causal network inference via group sparse regularization,” *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2628–2641, 2011.

[6] J. Mei and J. M. F. Moura, “Signal processing on graphs: Causal modeling of unstructured data,” *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2077–2092, 2017.

A Annex

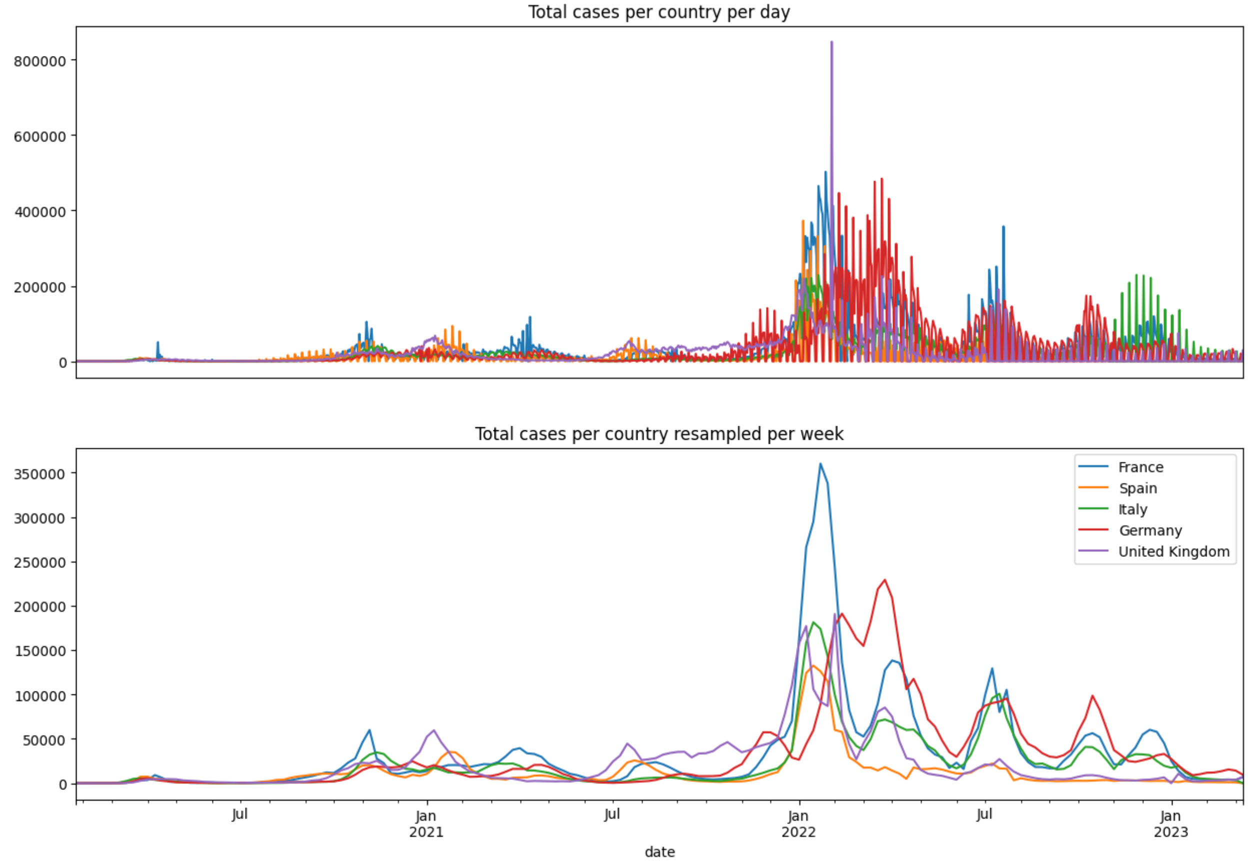

Figure 3: Covid cases for somes countries during time

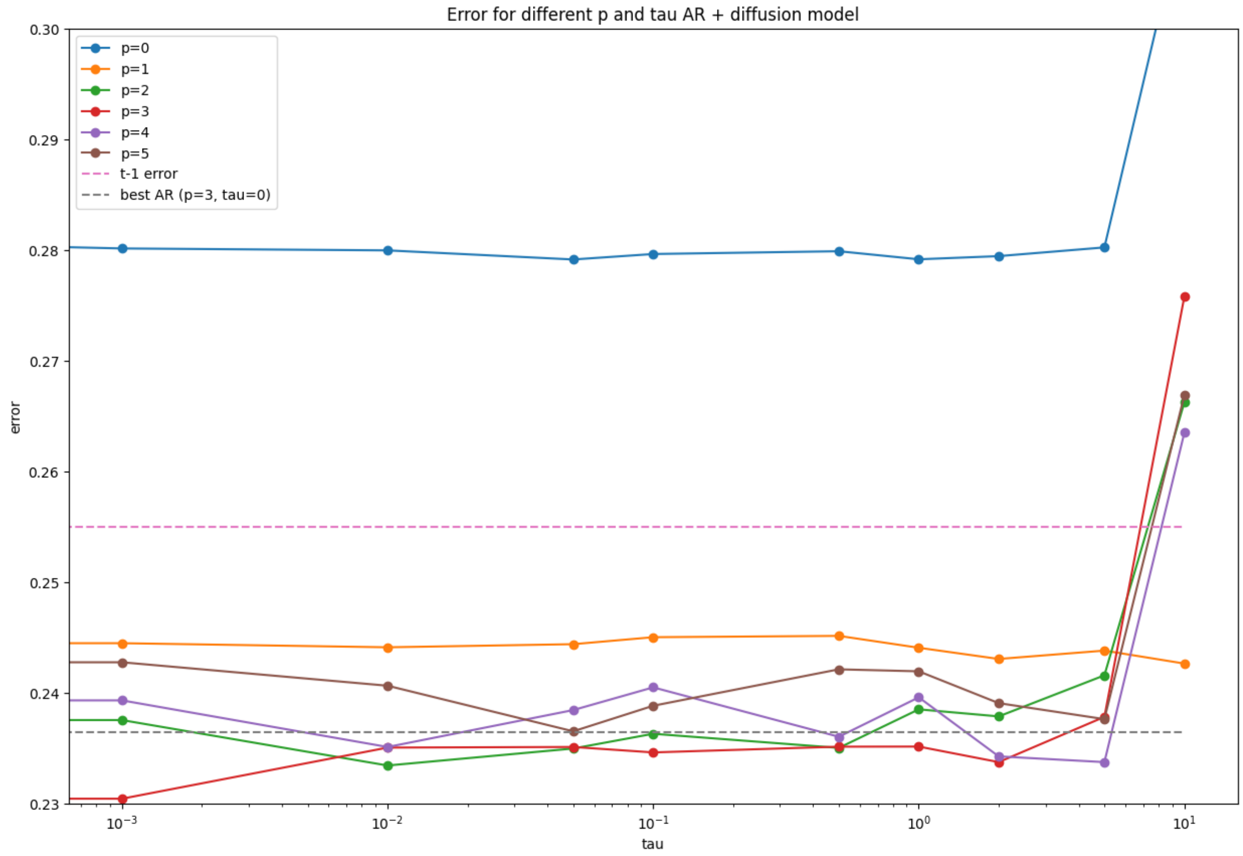

Figure 4: Error for different p and tau AR + diffusion model

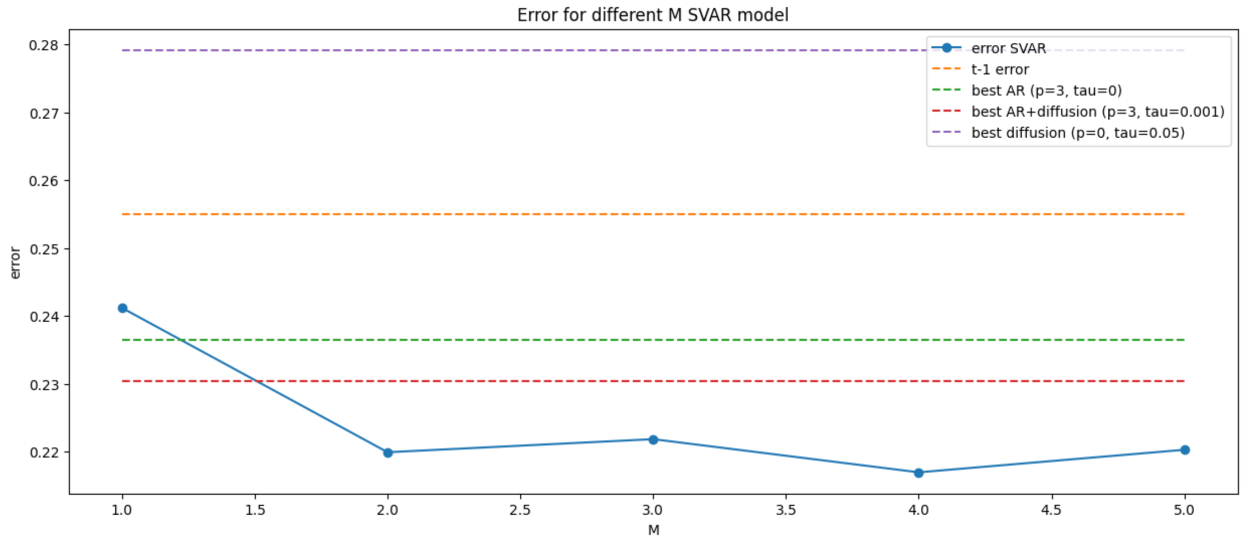

Figure 5: Error for different M - SVAR model