

An Explainable Estimation Model for Financial Time Series

Osamu Takahashi¹, Annie Cuyt², Anthony O'Hare², Wen-shin Lee², and
Massimiliano Ferrara^{1,3}

¹ Department of Law, Economics, and Human Sciences
Mediterranean University of Reggio Calabria, Italy
{osamu.takahashi, massimiliano.ferrara}@unirc.it

² Division of Computing Science and Mathematics
University of Stirling, United Kingdom
{annie.cuyt, anthony.ohare, wen-shin.lee}@stir.ac.uk

³ ICRIOS - Bocconi University, Milan, Italy

Abstract. Recent developments in artificial intelligence have led to powerful machine learning approaches to model fluctuating dynamics in financial time series, even in the presence of non-stationary behavior. However, such methods often suffer from limited interpretability while uncertainty, arising from complex data fluctuations, remains difficult to quantify.

We present a method that decomposes financial data into different frequency components and trends, thereby enabling explicit identification of cyclical patterns and long-term trends. Concurrently, noise components are separated from signal components. Its effectiveness and its high predictive accuracy have been demonstrated in the interpretation of some financial time series.

1 Introduction and Motivation

A main challenge in time frequency analysis is the development of explainable models for fluctuating data without relying on rigid modeling assumptions. Recent advances in artificial intelligence have led to the widespread adoption of machine learning methods, such as recurrent neural networks, due to their strong predictive performance [2]. However, their complexity often limits interpretability and transparency in decision-making.

Classical autoregressive approaches, including Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, have been combined with machine learning techniques [2]. Nevertheless, their reliability is constrained by a limited ability to distinguish structural components from noise and a strong dependence on lagged dependent variables [5].

On the other hand, signal processing techniques have been widely applied to financial data due to their ability to decompose cyclical components or long-term trends while separating noise. Among them, the Discrete Fourier Transform

(DFT) and Singular Spectrum Analysis (SSA) are frequently used. However, Fourier-based approaches assume periodicity [1] while SSA does not retrieve explicit parameters such as frequencies and damping factors that are essential for an estimation model [5]. To address these issues, we present a framework for financial time series, based on the Matrix Pencil (MP) method [3], a modern version of the classical Prony method [4].

We integrate recent developments in exponential analysis to enhance interpretability and demonstrate the effectiveness of the approach on real financial datasets in terms of reconstruction and predictive accuracy. The proposed framework emphasizes interpretability by explicitly linking the extracted exponential components to financial concepts such as trend, seasonality, and transient effects.

2 Methodology

2.1 Time series modeling with the MP method

We briefly summarize the model behind the MP method. The observed time series data y_j at time t_j are decomposed into two parts: a structural part represented by a sum of complex exponential functions, denoted by $f(t_j)$, and residuals, denoted by $r(t_j)$:

$$y_j = f(t_j) + r(t_j). \quad (1)$$

The time points $t_j = j\Delta$, $j = 0, \dots, N-1$, for a total of N observations, are taken equidistantly where Δ denotes the sampling interval. In typical financial time series applications, Δ is set to unity without loss of generality.

The MP method extracts and decomposes structural components, such as cyclical components and long-term trends, as

$$f(t_j) = \sum_{i=1}^{\mu} \alpha_i \exp(\phi_i t_j), \quad (2)$$

where $\phi_i \in \mathbb{C}$ characterize cyclical components when they have non-zero imaginary parts and trend components when they are real. The cyclical components can be damped or can grow. We call the coefficients $\alpha_i \in \mathbb{C}$ the amplitudes. The model is explainable because each component corresponds to an (α_i, ϕ_i) pair, thus allowing interpretation in terms of trend, seasonality, and transient effects.

As in SSA, the MP method employs the Singular Value Decomposition (SVD) of a Hankel matrix constructed from the time series. A Hankel matrix is a rectangular matrix with constant values along anti-diagonals. The numerical rank of the Hankel matrix, determined using a data-driven criterion based on the decay of its singular values, determines the number of terms μ .

Subsequently, the MP method differs substantially from SSA. MP identifies the exponential terms in (2) explicitly, through the solution of a Hankel structured generalized eigenvalue problem.

2.2 Hankel embedding and Singular Value Decomposition (SVD)

The time series is embedded into a Hankel matrix $H \in \mathbb{R}^{L \times M}$

$$H = \begin{pmatrix} y_0 & y_1 & \cdots & y_{M-1} \\ y_1 & y_2 & \cdots & y_M \\ \vdots & \vdots & \ddots & \vdots \\ y_{L-1} & y_L & \cdots & y_{N-1} \end{pmatrix}, \quad (3)$$

where $L = N - M + 1$ and $L \geq M$.

Computing the singular value decomposition gives

$$H = USW^T, \quad (4)$$

where $S = \text{diag}(s_1, \dots, s_M)$ contains the singular values ordered according to decreasing magnitude. Small singular values are typically associated with noise, and a threshold τ is used to determine the numerical rank μ of H . The threshold τ can be selected based on the relative decay of singular values.

2.3 Identifying the exponential terms

We define $\Phi_i := \exp(\phi_i \Delta)$. The unitary matrix $U \in \mathbb{C}^{L \times L}$ in (4) contains the left singular vectors:

$$U = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,L} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ u_{L,1} & u_{L,2} & \cdots & u_{L,L} \end{pmatrix}. \quad (5)$$

Applying the threshold τ to the singular values s_i , a truncated unitary matrix $\tilde{U} \in \mathbb{C}^{L \times \mu}$ is constructed by retaining the μ dominant left singular vectors:

$$\tilde{U} = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,\mu} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,\mu} \\ \vdots & \vdots & \ddots & \vdots \\ u_{L,1} & u_{L,2} & \cdots & u_{L,\mu} \end{pmatrix}. \quad (6)$$

Next, the matrices $\tilde{U}^{(0)}, \tilde{U}^{(1)} \in \mathbb{C}^{(L-1) \times \mu}$ are formed by removing the last and first rows, respectively:

$$\tilde{U}^{(0)} = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,\mu} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,\mu} \\ \vdots & \vdots & \ddots & \vdots \\ u_{L-1,1} & u_{L-1,2} & \cdots & u_{L-1,\mu} \end{pmatrix}, \quad \tilde{U}^{(1)} = \begin{pmatrix} u_{2,1} & u_{2,2} & \cdots & u_{2,\mu} \\ u_{3,1} & u_{3,2} & \cdots & u_{3,\mu} \\ \vdots & \vdots & \ddots & \vdots \\ u_{L,1} & u_{L,2} & \cdots & u_{L,\mu} \end{pmatrix}. \quad (7)$$

The generalised eigenvalue Φ_i is now obtained by solving the generalised eigenvalue problem:

$$\tilde{U}^{(1)} w_i = \Phi_i \tilde{U}^{(0)} w_i. \quad (8)$$

From Φ_i the frequency parameter ϕ_i can be determined, subject to the condition

$$\max_i |\Im(\phi_i)\Delta| < \pi,$$

where we usually set $\Delta = 1$ in financial time series.

The amplitudes α_i in (2) are obtained from the interpolation conditions:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \Phi_1 & \Phi_2 & \cdots & \Phi_\mu \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1^{N-1} & \Phi_2^{N-1} & \cdots & \Phi_\mu^{N-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_\mu \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{pmatrix} \quad (9)$$

2.4 Forecast and residuals

The above procedure yields the frequency parameters ϕ_i and amplitudes α_i introduced in (2), from the observed data. These parameters are then used to forecast future values of the time series. Future values are predicted by extrapolating the exponential model (2) over a short period of time beyond the observed interval.

The prediction accuracy is defined as the relative error

$$\eta_j = \left| \frac{y_j - f(t_j)}{y_j} \right| = \left| \frac{r(t_j)}{y_j} \right|. \quad (10)$$

The residuals r_j are non-structural components representing uncertainty in the financial time series model.

3 Result

Experiments using the MP method on real financial datasets demonstrate its ability to capture both seasonal fluctuations and long-term trends in financial transactions.

In this short paper, we focus on illustrating the interpretability of the proposed method rather than providing an exhaustive empirical evaluation. The extracted components reveal that certain cycles are more significant than others, such as quarterly cycles, which is consistent with the fact that corporate budget plans are typically reviewed on a quarterly basis.

In our experiments the relative deviation between the reconstructed time series and the original time series was below 3% on average, with the error distribution exhibiting positive skewness. Similarly, the average prediction error between the MP-based forecasting model and the observed data remains below 5%, also showing a positively skewed error distribution. These error levels are considered significantly low from a practical statistical perspective in financial time series analysis.

Moreover, the proposed method captures the magnitude and directional asymmetry of cycle fluctuations, providing useful information for financial risk assessment. From a methodological perspective, the MP approach does not rely on recursive updates of lagged variables, which are typically used in conventional autoregressive models such as ARIMA, and related time series frameworks. In particular, it can distinguish cases in which either peak or trough behaviour is more informative than the counterpart.

4 Discussion and Outlook

The new approach successfully expresses the significant seasonal fluctuations in the observed time series and demonstrates high accuracy in predicting the future. Detailed formulation of the procedure, model selection criteria, and experimental results are described in a forthcoming paper.

The uncertainty dynamics of the non-structured part can be further analyzed through an ARCH-type method applied to the extracted residuals, enabling the assessment of time-varying volatility. In addition, some machine learning-based model selection strategy may be incorporated.

Acknowledgement

This research received funding from the European Union Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement 101008231 (EXPOWER).

References

1. Cuyt, A., Tsai, M.N., Verhoye, M., s. Lee, W.: Faint and clustered components in exponential analysis. *Applied Mathematics and Computation* **327**, 93–103 (2018). <https://doi.org/10.1016/j.amc.2017.11.007>
2. Dixon, M.F., Halperin, I., Bilokon, P.: *Machine Learning in Finance*. Springer International Publishing, Cham (2020). <https://doi.org/10.1007/978-3-030-41068-1>
3. Hua, Y., Sarkar, T.K.: Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **38**(5), 814–824 (1990). <https://doi.org/10.1109/29.56027>
4. Prony, G.R.: Essai expérimental et analytique. *Journal de l'École Polytechnique* **2**, 24–76 (1795)
5. Thomakos, D.D., Wang, T., Wille, L.T.: Modeling daily realized futures volatility with singular spectrum analysis. *Physica A* **312**(3–4), 505–519 (2002). [https://doi.org/10.1016/S0378-4371\(02\)00845-2](https://doi.org/10.1016/S0378-4371(02)00845-2)