ENHANCING STOCK PRICE DATA ANALYSIS THROUGH VARIANTS OF PRINCIPAL COMPONENT ANALYSIS
(Penambahbaikan Analisis Data Harga Saham Menerusi Varian-varian Analisis Komponen Prinsipal)

SEOW TECK XIANG, DHARINI PATHMANATHAN* & KHOO TZUNG HSUEN

ABSTRACT

This work investigates and identifies suitable dimensionality reduction approaches based on variants of principal component analysis (PCA) for various transformations of stock price data. The classical PCA, dynamic principal component analysis (DPCA) and generalised dynamic principal component analysis (GDPCA) were applied to the closing prices, simple returns and log of returns of the top 100 holdings of Standard & Poor's 500 (S&P500) from year 2020 to year 2023. The S&P 500 is a stock market index that tracks the stock performance of 500 large-cap U.S. companies. The performances of the aforementioned variants of PCA on these data for different timeframes were compared. Results showed that GDPCA works best for non-stationary time series data such as the closing prices and DPCA works best for stationary time series data such as the simple returns and the log of returns. The results obtained from the empirical analysis was further supported by simulation studies that follow, hence GDPCA and DPCA could be among the most appropriate dimensionality reduction approaches for non-stationary and stationary time series data respectively.

Keywords: stock market; non-stationary time series; geometric Brownian Motion; GJR-GARCH; GARCH

ABSTRAK

Kajian ini menyelidik dan mengenalpasti pendekatan-pendekatan pengurangan dimensi yang sesuai berpandukan varian-varian analisis komponen prinsipal untuk pelbagai penjelmaan data harga saham. Analisis komponen prinsipal statik, analisis komponen prinsipal dinamik dan analisis komponen prinsipal dinamik teritlak telah dikenakan pada harga tutup, pulangan mudah dan pulangan log harian untuk 100 saham terbesar yang tersenarai dalam Standard & Poor's 500 (S&P500) dari tahun 2020 ke tahun 2023. S&P500 merupakan indeks pasaran saham yang mengesan prestasi saham bagi 500 syarikat dengan modal pasaran yang besar di Amerika Syarikat. Prestasi varian-varian tersebut atas data yang mempunyai jangka masa berbeza juga telah dibandingkan. Keputusan menunjukkan bahawa analisis komponen prinsipal dinamik teritlak merupakan varian yang paling sesuai untuk siri-siri masa tidak pegun seperti data harga tutup saham manakala analisis komponen prinsipal dinamik adalah paling sesuai untuk siri-siri masa pegun seperti pulangan mudah dan pulangan log. Keputusan daripada simulasi-simulasi yang dijalankan seterusnya juga menyokong kesimpulan yang telah diperoleh daripada analisis empirikal, justeru analisis komponen prinsipal dinamik teritlak ialah antara pendekatan-pendekatan pengurangan dimensi siri masa tidak pegun yang paling berkesan manakala analisis komponen prinsipal dinamik ialah antara pendekatan-pendekatan pengurangan dimensi siri masa pegun yang paling sesuai.

Kata kunci: pasaran saham; siri masa tidak pegun; gerakan Brownan geometri; GJR-GARCH; GARCH

References

Ashraf M., Anowar F., Setu J.H., Chowdhury A.I., Ahmed E., Islam A. & Al-Mamun A. 2023. A survey on dimensionality reduction techniques for time-series data. *IEEE Access* **11**: 42909-42923.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3): 307-327.

Brillinger D.R. 1981. *Time series: Data analysis and theory (Holden-Day series in time series analysis).* San Francisco, CA: Holden Day.

Bruna M.G., Loprevite S., Raucci D., Ricca B. & Rupo D. 2022. Investigating the marginal impact of ESG results on corporate financial performance. *Finance Research Letters* **47**: 102828.

Donadelli M. & Paradiso A. 2014. Is there heterogeneity in financial integration dynamics? Evidence from country and industry emerging market equity indexes. *Journal of International Financial Markets, Institutions and Money* **32**: 184-218.

Elliott G., Granger C.W.J. & Timmermann A.G. 2006. *Handbook of Economic Forecasting.* Amsterdam, The Netherlands: North-Holland.

Ghalanos A. 2022. rugarch: Univariate GARCH models. R package version 1.4-9.

Glosten L.R., Jagannathan R. & Runkle D.E. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* **48**(5): 1779-1801.

Hormann S. & Kidzinski L. 2022. freqdom: Frequency Domain Based Analysis: Dynamic PCA. R package version 2.0.3.

Imai J. & Tan K.S. 2006. A general dimension reduction technique for derivative pricing. *Journal of Computational Finance* **10**(2): 129-155.

Jolliffe I.T. 2002. *Principal Component Analysis*. 2nd Ed. New York City, NY: Springer-Verlag.

Kambouroudis D.S., McMillan D.G. & Tsakou K. 2016. Forecasting stock return volatility: A comparison of GARCH, implied volatility, and realized volatility models. *Journal of Futures Markets* **36**(12): 1127-1163.

Khoo T.H., Dabo I.M., Pathmanathan D. & Dabo-Niang S. 2024. Generalized functional dynamic principal component analysis. *arXiv preprint arXiv*: 2407.16024.

Ku W., Storer R.H. & Georgakis C. 1995. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **30**(1): 179-196.

Kumar S. 2022. Effective hedging strategy for US Treasury bond portfolio using principal component analysis. *Academy of Accounting and Financial Studies Journal* **26**(1): 1-17.

Liu C. 2022. Risk prediction of digital transformation of manufacturing supply chain based on principal component analysis and backpropagation artificial neural network. *Alexandria Engineering Journal* **61**(1): 775-784.

Lòpez de Prado M.M. 2020. *Machine Learning for Asset Managers.* Cambridge, UK: Cambridge University Press.

Ma F., Guo Y., Chevallier J. & Huang D. 2022. Macroeconomic attention, economic policy uncertainty, and stock volatility predictability. *International Review of Financial Analysis* **84**: 102339.

Mader H.M., Coles S.G., Connor C.B. & Connor L.J. 2006. *Statistics in Volcanology*. London, UK: The Geological Society of London.

Mancino M.E. & Renò R. 2005. Dynamic principal component analysis of multivariate volatility via Fourier analysis. *Applied Mathematical Finance* **12**(2): 187-199.

Marathe R.R. & Ryan S.M. 2005. On the validity of the geometric Brownian motion assumption. *The Engineering Economist* **50**(2): 159-192.

McDonald R.L. 2013. *Derivatives Markets*. 3rd Ed. Upper Saddle River, NJ: Pearson Education.

Nugroho D.B., Kurniawati D., Panjaitan L.P., Kholil Z., Susanto B. & Sasongko L.R. 2019. Empirical performance of GARCH, GARCH-M, GJR-GARCH and log-GARCH models for returns volatility. *Journal of Physics: Conference Series* **1307**(1): 012003.

Pearson K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**(11): 559-572.

Peña D., Smucler E. & Yohai V.J. 2020. gdpc: An R Package for Generalized Dynamic Principal Components. *Journal of Statistical Software, Code Snippets* **92**(2): 1-23.

Peña D. & Yohai V.J. 2016. Generalized dynamic principal components. *Journal of the American Statistical Association* **111**(515): 1121-1131.

Reddy K. & Clinton V. 2016. Simulating stock prices using geometric Brownian motion: Evidence from Australian companies. *Australasian Accounting, Business and Finance Journal* **10**(3): 23-47.

Ryan J.A., Ulrich J.M., Smith E.B., Thielen W., Teetor P. & Bronder S. 2022. quantmod: Quantitative Financial Modelling Framework. R package version 0.4.20.

Sarıkoç M. & Celik M. 2024. PCA-ICA-LSTM: A hybrid deep learning model based on dimension reduction methods to predict S&P 500 index price. *Computational Economics*: 1-67.

Song Z., Gong X., Zhang C. & Yu C. 2023. Investor sentiment based on scaled PCA method: A powerful predictor of realized volatility in the Chinese stock market. *International Review of Economics & Finance* **83**: 528-545.

Wang X. 2006. On the effects of dimension reduction techniques on some high-dimensional problems in finance. *Operations Research* **54**(6): 1063-1078.

Wei J. & Huang J. 2012. An exotic long-term pattern in stock price dynamics. *PLoS One* **7**(12): e51666.

Ying X. 2019. An overview of overfitting and its solutions. *Journal of Physics: Conference Series* **1168**: 022022.

Zhang Y. & Wang Y. 2023. Forecasting crude oil futures market returns: A principal component analysis combination approach. *International Journal of Forecasting* **39**(2): 659-673.

Zhang Z. 2022. Research on stock price prediction based on PCA-LSTM model. *Academic Journal of Business & Management* **4**(3): 42-47.

Zhong X. & Enke D. 2017. Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications* **67**: 126-139.

Institute of Mathematical Sciences
Faculty of Science
Universiti Malaya
50603 Kuala Lumpur
Wilayah Persekutuan Kuala Lumpur, MALAYSIA
E-mail: s2025219@siswa.um.edu.my, dharini@um.edu.my[*], 17201135@siswa.um.edu.my

[*]Corresponding author