Analysis of the Correlation Between Chinese and American Stock Markets Based on VAR Model

Mengqi Shi*

Department of Mathematical Sciences, Tsinghua University, Beijing, China

* Corresponding Author Email: shimq21@mails.tsinghua.edu.cn

Abstract. This study investigates the correlation between Chinese and American stock markets during the post-pandemic era, employing a Vector Autoregression (VAR) model and analyzing data from the SSEC Index and S&P 500 Index spanning from April 9th, 2019, to April 3rd, 2024. The analysis reveals stationary behavior in both indexes and highlights a bi-directional influence between them, with the S&P 500 exerting a notably stronger impact on the SSEC. Impulse response analysis indicates that following a shock to the S&P 500, the SSEC responds after approximately 20 days, while a shock to the SSEC results in a lagged response of about 23 days in the S&P 500. Variance decomposition analysis underscores the dominance of endogenous shocks, with SSEC's shocks accounting for approximately 98% of its forecast error variance and S&P 500's shocks contributing to about 97% of its forecast error variance. The declining trend in forecast error variance contribution over time suggests an increasing certainty and accuracy in the model's predictions regarding the relationship between these markets. These insights provide valuable guidance for investors, policymakers, and researchers navigating the complexities of the post-pandemic financial landscape.

Keywords: Stock market correlation, VAR model, granger causality, impulse response function, variance decomposition.

1. Introduction

The interconnectedness of global financial markets has become increasingly apparent in today's interconnected world, with the relationship between the Chinese and American stock markets holding particular significance. As two of the world's largest economies and major players in global trade and finance, understanding the dynamics of their relationship is crucial. Changes in these stock markets can have far-reaching effects, impacting investor sentiment, economic policies, and financial stability worldwide. Moreover, in the wake of ongoing trade tensions and geopolitical dynamics between China and the United States, studying their stock market relationship can provide insights into broader economic and political trends, aiding in anticipating potential risks and opportunities in the global financial landscape.

Numerous studies have delved into the relationship between Chinese and American stock markets, yet many overlook quantitative model and timeliness analysis. For instance, the dynamic interdependence between the US and Asian markets has been explored by Dhanaraj, emphasizing empirical analysis [1]. However, the study's focus on Asian markets may lack specificity regarding the Chinese American relationship. Similarly, Fan investigates the dynamic linkages between China and international stock markets, but their analysis might overlook nuances specific to the US-China relationship [2]. Yang delves into market segmentation and information asymmetry in Chinese stock markets using VAR analysis, yet its narrow focus on information asymmetry may limit its broader applicability [3]. Additionally, Wen analyzes risk spillovers between oil and stock markets using VAR for VaR analysis, which although insightful, may not directly address the Chinese American dynamic [4]. Trung examines the spillover effects of US economic policy uncertainty on the global economy employing a global VAR approach, yet its focus on policy uncertainty may not fully capture the breadth of factors influencing the Chinese American relationship [5]. While these studies contribute valuable insights, their limitations underscore the need for a focused investigation into the specific dynamics between the Chinese and American stock markets, particularly in the post-pandemic era. This paper seeks to address these gaps by employing advanced econometric techniques

to unravel the complex interactions between these markets, offering timely and relevant insights for academia and practical applications in finance and economics.

Considering these limitations, this paper aims to contribute to the research by conducting a comprehensive analysis of the relationship between the Chinese and American stock markets, with a particular focus on the post-pandemic era. By utilizing high-quality and up-to-date data from the SSEC Index and S&P 500 Index, this study seeks to capture the full spectrum of market dynamics. Employing advanced econometric techniques, including Vector Autoregression (VAR) modeling, Granger causality testing, impulse response analysis, and variance decomposition, the paper aims to unravel the complex interactions between the two markets. The objective is to provide empirical evidence on the direction, strength, and persistence of spillover effects, causal relationships, and response patterns between Chinese and American stocks. Furthermore, this investigation seeks to identify potential implications for investors, policymakers, and market participants, offering insights into portfolio diversification, risk management, and strategic decision-making in a globalized financial landscape. The novelty of the study lies in its timeliness, as it aims to quantify the impact of recent events on the relationship between these markets, thus contributing valuable insights to both academic research and practical applications in finance and economics.

2. Data and Pre-processing

The paper mainly uses the data of SSEC Index and S&P 500 Index, representing both China and America stock markets. The data resource is Investing (https://cn.investing.com/), and the data includes the date, closing prices, opening prices and volume of both indexes. The time period is from April 9th 2019 to April 3rd 2024 and the time length is 5 years, giving better robustness and wider application of quantitative model. In pre-process, the returns of both indexes are calculated using the closing prices. Here are the statistical characters of returns of SSEC and S&P 500 (See Table 1).

Table 1. Statistical characters of returns

	standard deviation	average	maximum	minimum	skewness	kurtosis
SSEC	1.0389	0.0160	8.3711	-3.3664	0.9599	6.0383
S&P 500	1.3591	-0.0479	13.6158	-8.5779	1.1977	16.9389

3. Model Selection

The Vector Autoregression (VAR) model is pivotal in econometrics and time series analysis due to several reasons. Firstly, it facilitates the concurrent examination of multiple variables, capturing their dynamic interdependencies. This feature is critical for comprehending intricate systems like economics and finance, where variables interact extensively. VAR's consideration of these interrelationships offers a more holistic understanding compared to single-equation models. Secondly, the bivariate VAR model delves into the dynamic relationship between two variables, incorporating lagged values of both. This allows researchers to scrutinize the mutual influence between variables over time, shedding light on how changes in one affect the other and vice versa. Such analysis is invaluable for exploring closely linked pairs of variables, such as interest rates and inflation or exchange rates and trade balances.In a bivariate VAR model, each variable is expressed as a function of its own lagged values as well as lagged values of the other variable. The model is given as:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{p} \gamma_i X_{t-i} + \varepsilon_t \tag{1}$$

$$X_t = \delta + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{p} \theta_i Y_{t-i} + \eta_t \tag{2}$$

where Y_t and X_t are the two variables of interest at time t, α, β_i, γ_i, δ, φ_i and θ_i are coefficients to be estimated, p represents the lag order of the model, ε_t and η_t are error terms.

In summary, VAR models offer a robust framework for analyzing the dynamic interactions between multiple variables, while bivariate VAR specifically focuses on the relationship between two variables, making it a versatile and widely used tool in empirical research.

4. Results

By applying the bivariate VAR model on SSEC Index and S&P 500 Index, this part aims to reveal the statistical results of calculation and influence of one index on another one, illustrating the deeper interaction of Chinese and American stock markets.

4.1. Augmented Dickey-Fuller Test

The Augmented Dickey-Fuller (ADF) test is a pivotal tool in econometrics for assessing the stationarity of time series data. It determines whether a series has a unit root. The null hypothesis of the ADF test indicates the presence of a unit root and non-stationarity in the time series. If the null hypothesis is rejected, it suggests that the time series is stationary. The ADF test results are presented in Table 2.

Table 2. ADF test results

ADF test	ADF Statistic	p-value
SSEC	-16.046	0.000
S&P 500	-10.506	0.000

Based on the results in Table 2 for both series (SSEC and sp500), the ADF statistic is substantially lower than the critical values at any confidence level, with corresponding p-values of 0.000. This indicates rejection of the null hypothesis, rejecting the presence of a unit root, and thus, both the series are stationary.

4.2. Freedom of Lag Selection

The freedom of lag selection in the Augmented Dickey-Fuller (ADF) test refers to the ability to choose the lag length p in the regression equation. This freedom is crucial as it impacts the test's accuracy and reliability in assessing the stationarity of a time series. Several criteria are comprehensively employed to determine the optimal lag length. By allowing flexibility in lag selection, the ADF test accommodates different data characteristics and ensures robust analysis of time series data. The lag order selection processes are based on AIC, BIC, FPE, HQIC criteria. Based on the Table 3, the criteria for lag order selection indicate that the lag order of 2 is highlighted as the optimal choice. This selection is consistent across multiple criteria, including AIC and FPE, which are commonly used for lag order determination. A lag order of 2 minimizes these information criteria, suggesting it provides the best balance between model fit and complexity.

Table 3. Lag Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	1.091	1.097	2.978	1.093
1	0.6178	0.6360	1.855	0.6245
2	0.5559	0.5863*	1.744	0.5671*
3	0.5586	0.6011	1.748	0.5743
4	0.5595	0.6141	1.750	0.5797
5	0.5614	0.6282	1.753	0.5860
6	0.5649	0.6437	1.759	0.5940
7	0.5496	0.6407	1.733	0.5832
8	0.5429*	0.6460	1.721*	0.5810

4.3. Granger Causality Test

After having the optimal lag selection, the paper uses the statistical method to determine if one time series variable can be used to predict another. The Granger causality test is typically implemented within a VAR framework. The Granger causality test is used to determine whether there is a causal relationship between two time series. In these results, it can be observed that the statistics and p-values of the F-test based on SSR, chi-squared test based on SSR, likelihood ratio test, and parameter F-test for different lag orders. From Table 4 and Table 5, at each lag order, the p-values of the F-test based on SSR, chi-squared test based on SSR, likelihood ratio test, and parameter F-test are all less than 0.05. This implies the rejection of the null hypothesis. Therefore, S&P500 Granger causes SSEC, and this causal relationship remains strong even with 2 lags; SSEC Granger may cause SP500, but the evidence is weaker with 2 lags.

Table 4. Granger Causality: SP500→SSEC

Number of lags(no zero)	Statistical results
2	ssr based F test: F=2.9230, p=0.0540, df_denom=1815, df_num=2 ssr based chi2 test: chi2=5.8622, p=0.0533, df=2 likelihood ratio test: chi2=5.8528, p=0.0536, df=2 parameter F test: F=2.9230, p=0.0540, df_denom=1815, df_num=2

Table 5. Granger Causality: SSEC→SP500

Number of lags(no zero)	Statistical results
2	ssr based F test: F=11.2011 , p=0.0000, df_denom=1815, df_num=2 ssr based chi2 test: chi2=22.4640 , p=0.0000, df=2 likelihood ratio test: chi2=22.3265 , p=0.0000, df=2 parameter F test: F=11.2011 , p=0.0000, df_denom=1815, df_num=2

4.4. Impulse Response Function

The impulse response function (IRF) is a fundamental concept in VAR models, illustrating the dynamic response of variables to a shock in the system. In a VAR model, the IRF quantifies the impact of an exogenous shock to one variable on all variables in the system over time. By analyzing the impulse response functions, researchers can understand the dynamic interactions between variables in a VAR model and assess the persistence and magnitude of shocks over time. The corresponding impulse responses after orthogonalization are presented in Fig. 1.

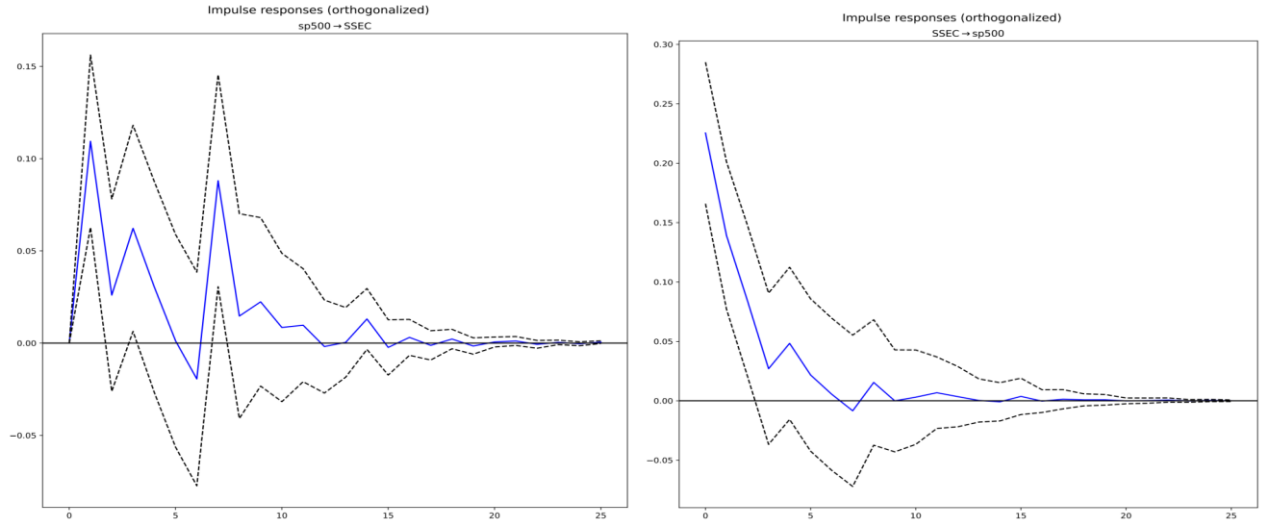


Fig. 1 Impulse responses

For each directional impulse shock, there are three curves: two black curves representing confidence intervals (typically at a 95% confidence level), and a blue curve representing the average response path. The intersection with the horizontal axis indicates the time point (unit: day) after the shock, and the vertical axis value represents the degree of change relative to the baseline. This time point serves to measure the lag effects of the dynamic relationship between variables.

For the 'SP500→SSEC' plot, it indicates the response of SSEC when SP500 experiences a shock. The blue curve similarly shows how SSEC responds relative to the shock, which is 20 days. The 'SSEC→SP500' plot illustrates the response of SSEC when SP500 is subject to a shock, and the lagged time is about 23 days.

4.5. Variance Decomposition Analysis

The variance decomposition analysis is a crucial aspect of VAR models, offering insights into the proportion of forecast error variance in each variable attributable to its own shocks and the shocks of other variables in the system. This analysis helps identify the main drivers of forecast uncertainty in a VAR model and assesses the relative importance of each variable's shocks in explaining the variability of other variables over time. The variance decomposition results are calculated as follows in Table 6 and Fig. 2.

Table 6. Variance decomposition results

	FEVD for SSEC			FEVD for S&P500	
	SSEC	S&P500		SSEC	S&P500
0	1.000000	0.000000	0	0.029843	0.970157
1	0.990670	0.009330	1	0.038639	0.961361
2	0.991392	0.008608	2	0.040406	0.959594
3	0.989276	0.010724	3	0.040593	0.959407
4	0.988876	0.011124	4	0.041257	0.958743
5	0.988982	0.011018	5	0.041487	0.958513
6	0.988821	0.011179	6	0.041498	0.958502
7	0.984111	0.015889	7	0.040946	0.959054
8	0.984016	0.015984	8	0.041039	0.958961
9	0.983729	0.016271	9	0.041027	0.958973

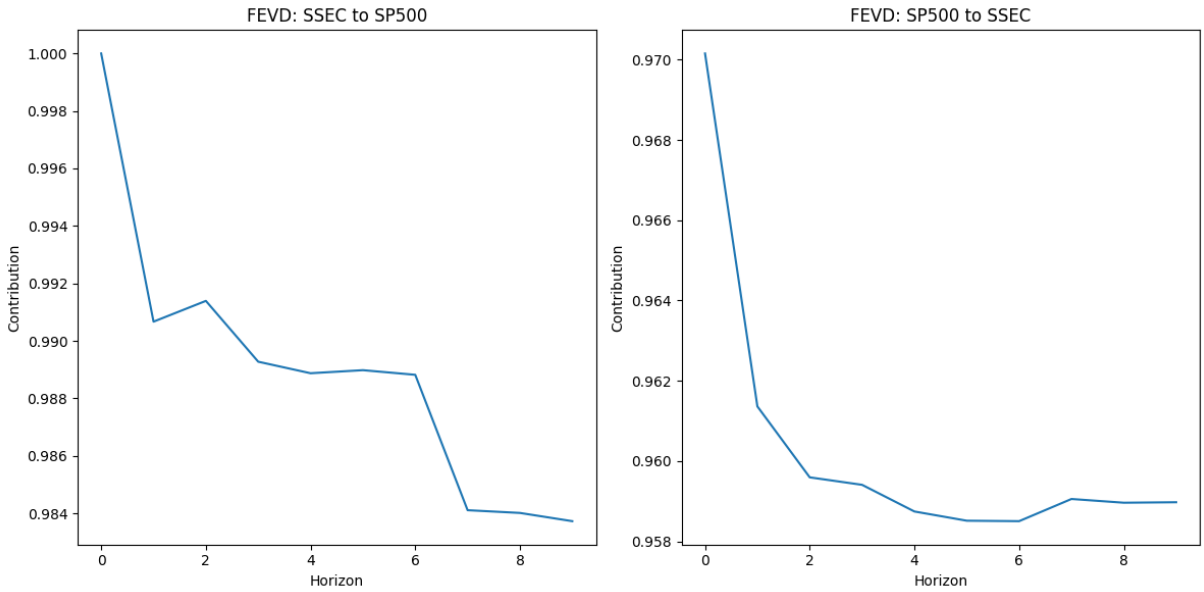


Fig. 2 Forecast Error Variance Decomposition

The x-axis in represents the forecast horizon, indicating the time steps to predict. The y-axis represents the contribution of each variable to the forecast error variance, reflecting the degree to which each variable contributes to the model's prediction accuracy. For 'SP500→SSEC' and 'SSEC→SP500' plots, the downward trend on the y-axis indicates that over the forecast horizon, the contribution of each variable to the forecast error variance gradually decreases over time. This declining trend may suggest that as time progresses, the model's predictions regarding the relationship between these variables become more certain and increasingly accurate, leading to a reduction in their contribution to the forecast error.

5. Conclusion

In summary, this study provides a comprehensive analysis of the relationship between the Chinese and American stock markets, utilizing data from the SSEC Index and S&P 500 Index over a 5-year period. Through the application of the Vector Autoregression (VAR) model and various econometric techniques, significant findings regarding the dynamic interactions between these markets have been revealed. The results indicate a bi-directional influence between the markets, with stronger evidence of S&P 500 impacting SSEC than vice versa. Impulse response analysis illustrates the time lag and magnitude of responses to shocks in either market, while variance decomposition analysis dissects the forecast error variance, elucidating the relative contributions of each market to predictive accuracy. These insights enhance the understanding of cross-market dynamics, aiding investors, policymakers, and researchers in strategic decision-making under the post-pandemic background.

However, this study has certain limitations. Firstly, the analysis relies on historical data and may not fully capture future market dynamics, particularly in rapidly evolving economic environments. Secondly, while advanced econometric techniques have been employed, the complexity of market interactions may require further refinement and validation through alternative methodologies. Despite these limitations, the findings of this study contribute valuable insights to the literature on stock market relationships and offer practical implications for stakeholders in the global financial landscape.

References

[1] Dhanaraj, S., Gopalaswamy, A. K., Babu M, S. Dynamic interdependence between US and Asian markets: an empirical study. Journal of Financial Economic Policy, 2013, 5(2): 220-237.

[2] Fan, K., Lu, Z., Wang, S. Dynamic linkages between the China and international stock markets. Asia-Pacific Financial Markets, 2009, 16: 211-230.

[3] Yang, J. Market segmentation and information asymmetry in Chinese stock markets: A VAR analysis. Financial Review, 2003, 38(4): 591-609.

[4] Wen, D., Wang, G. J., Ma, C., Wang, Y. Risk spillovers between oil and stock markets: A VAR for VaR analysis. Energy Economics, 2019, 80: 524-535.

[5] Trung, N. B. The spillover effects of US economic policy uncertainty on the global economy: A global VAR approach. The North American Journal of Economics and Finance, 2019, 48: 90-110.