ISSN: 2688-9323

# Capturing the Long Memory Attribute of Unemployment Rate Using the ARFIMA-EGARCH Model

Mengzhu Gao

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Anhui 233030, China

## Abstract

**Many previous analysis methods for the time series of unemployment rates assumed that the correlation decay rate of the series was fast, even exhibiting exponential decay. However, due to long-term dependencies in time series data, this assumption is violated in most cases. To fill this research gap, this article utilizes the autoregressive fractional integral moving average (ARFIMA GARCH) model with generalized autoregressive conditional heteroscedasticity to capture long memory features in the unemployment rate time series. Empirical analysis was conducted on the unemployment rate data of Hong Kong over the past 40 years, and the difference order of the model ($d = 0.396$) was used to confirm the existence of long memory attributes in the unemployment rate time series data.**

## Keywords

**Long Memory; ARFIMA-GARCH Model; Unemployment Rate.**

## 1. Introduction

In the past few decades, with the rapid development of China's economy, the phenomenon of unemployment has also become common. In the field of economics, a situation where a person is willing and able to work for compensation within a certain age range, but has not yet found a job, is considered unemployment. The obstruction of economic growth is related to the supply and demand of labor and other factor markets, and unemployment can have many negative effects. On a social level, unemployment poses a threat to the stability of families as social and economic units. Without income or income loss, the family's demands and needs cannot be met, and family relationships will be damaged as a result. Research has shown that the trauma caused by dismissal is no less than the death of relatives and friends or academic failure. Personal relationships outside of the family are also severely affected by unemployment, which may lead to a loss of self-esteem and confidence; The adverse effects of the unemployment economy are also significant, and the "three carriages" that drive the development of the national economy are investment, consumption, and exports. Unemployment means a decrease in residents' consumption levels, which will certainly affect economic development. When the unemployment rate rises, the products and services that could have been produced by unemployed workers in the economy are lost.

In previous studies, it was always assumed that the sequence correlation between time observation data would rapidly decrease or even exponentially decrease. However, some time series autocorrelation functions converge slowly and have long-term dependencies. Ignoring this feature may lead to incorrect inference. Therefore, this article applies the ARFIMA-EGARCH model to the time series analysis of Hong Kong's unemployment rate from 1982 to 2022. Firstly, the existence of long-term memory features in unemployment rate time series data was determined through unit root testing and the establishment of the ARFIMA model. Secondly, the existence of the ARCH effect was determined using the LM Arch test. Finally, the ARFIMA-

ISSN: 2688-9323

EGARCH model was applied for unemployment rate time series analysis. Empirical analysis is the second contribution of this study.

## 2. Literature Review

In recent years, there have been many studies on unemployment. Cai and Du concluded that the reason why economic growth did not bring about corresponding employment growth was mainly due to the limitations of counter-cyclical macroeconomic policies, and proposed that economic policies should focus on improving the labor mechanism, strengthening the construction of vocational training system, and promoting the development of high-employment industries[1]. Zeng and Yu based on the hypothesis of variable parameters, constructed a state space model including the process of natural unemployment rate change and the relationship between Phillips curve, and estimated the natural unemployment rate curve over time from 1992 to 2004 by Kalman filtering method. They found that China has a rising natural unemployment rate, and concluded that the most important policy goal of unemployment control should be to reduce the natural unemployment rate[2]. Su et al. found that delaying retirement can reduce China's unemployment rate. The timing of delaying the retirement age will have different effects on the unemployment rate. Carefully choosing the timing of delaying the start of retirement is of great significance for stabilizing employment[3]. Based on aggregated historical data, time series analysis helps to understand, explain, and predict trends in unemployment rates at an overall level. Previous research on unemployment time series data is shown in Table 1.

**Table 1.** Summary of time series analysis of unemployment rate in the past two decades

| Authors | Study region | Study Period | Temporal Unit | Method | Reseach Purpose |
|---|---|---|---|---|---|
| Han and An (2007) | Shenzhen | 1992-2006 | Year | VAR | Study the Impact of Minimum Wage on Unemployment and Labor Supply |
| Zhu et al. (2019) | Globe | 1992-2015 | Quarter | GVAR | Impact of China's New Normal and US Trade Protection on World Unemployment |
| Zhang and Jiang (2018) | China | 1996-2015 | Year | SVAR | Interaction between Unemployment Insurance and Employment |
| Li (2010) | China | 1978-2007 | Year | Cointegration model | Study on the relationship between urban unemployment rate and GDP growth rate, import and export |
| Yang (2013) | China | 1990-2010 | Year | VAR | Study on the Influence of Rural Floating Population on Urban Registered Unemployed Population |
| Zhang (2007) | China | 1978-1991 | Year | Granger causality test | On the Relationship between Urban - rural Income Gap and Urban Unemployment in China |
| Ding (2008) | kwangtung | 1978-2005 | Year | VAR | To study the interaction between employment level and resident income distribution, employment structure and resident income distribution |
| Xu and Wang (2009) | China | 1978-2007 | Year | EG test | Study on the Relationship between Registered Unemployment and Surveyed Unemployment |
| Wang (2021) | Shandong | 1980-2019 | Year | ARIMA | Predict urban registered unemployment rate |
| Yuan (2021) | China | 1999-2019 | Year | ARIMA | Projected unemployment rate |

ISSN: 2688-9323

Among the methods used in the above literature, the AR family model is widely used. Han and An used VAR model and annual data from 1992 to 2006 in Shenzhen to study the impact of minimum wage standards on unemployment and labor supply[4]; Zhu et al. examined the impact of China's new normal economy and US trade protection on unemployment in various countries around the world based on the GVAR model and national quarterly data from 1992 to 2015[5]; Zhang and Jiang examined the interaction between unemployment insurance and employment status based on the SVAR model and using national annual data from 1996 to 2011[6]; Li studied the relationship between urban unemployment rate and GDP growth rate, imports, and exports based on cointegration model, variance decomposition, and national annual data from 1978 to 2007[7]; Yang studied the impact of rural migrant population on urban registered unemployed population based on VAR model and national annual data from 1990 to 2010[8]; Zhang explored the relationship between urban-rural income gap and urban unemployment in China based on Granger causality test and national annual data from 1978 to 1991[9]; Ding used the VAR model and annual data from 1978 to 2005 in Guangdong to study the interrelationships between employment level and residents' income distribution, as well as the industrial structure of employment and residents' income distribution[10]; Xu and Wang studied the relationship between registered unemployment and surveyed unemployment based on EG test and national annual data from 1978 to 2007[11]; Wang used the ARIMA model and annual data from 1980 to 2019 in Shandong to predict the urban registered unemployment rate[12]; Yuan uses the ARIMA model and national annual data from 1999 to 2019 to predict future unemployment rates[13].

In these studies, it is assumed that the sequence correlation between time observation data will rapidly decrease or even exponentially decrease. However, some time series autocorrelation functions converge slowly. Li and Wei found that the stock market return series did not exhibit significant long-term memory throughout the entire sample interval. However, when extreme events occur, the return series exhibits significant correlation, reflecting the time-varying characteristics of the stock market's long-term memory. At this time, the analysis of historical data information can achieve the goal of avoiding extreme risks[14]; Liu and Qu used eight estimation methods from non parametric and semi parametric perspectives to empirically compare the long memory characteristics of the returns and volatility of major futures prices in China, represented by Shanghai copper, Shanghai gold, rubber, and Liandou. They found that overall, the long memory characteristics of futures price returns were not significant, while volatility had a significant long memory structure, This indicates that unpredictable information has a long-term impact on the volatility of China's futures market[15].

Although there are many analytical methods for studying the field of unemployment, there is little consideration given to long memory characteristics and heteroscedasticity issues in existing model frameworks. The contribution of this study is to establish a model that takes into account long memory characteristics and heteroscedasticity issues in traditional analysis (ARFIMA-EGARCH).

## 3. Methodology

The ARFIMA(p,d,q) model is defined as:

$$\phi(B)(1-B)^d(X_t-\mu_t)=\theta(B)\varepsilon_t \tag{1}$$

$$\mu_t=\mu_0+\sum_{i=1}^{m}\delta_i Y_{it} \tag{2}$$

ISSN: 2688-9323

Where $\varepsilon_t$ is white noise with zero mean and constant standard deviation; $X_{it}$ are the time-varying exogenous factors; $\delta_i$ and $\mu_0$ are vectors of coefficients and constants, respectively.

The $p$ parameter defines the polynomial of autoregressive coefficients in the delay operator $B$ : $\phi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p$ ; The $q$ parameter defines the polynomial of moving average coefficients in the delay operator $B:\theta(B)=1-\vartheta_1 B-\vartheta_2 B^2-\cdots-\vartheta_q B^q$.

$d$ in different regions indicates different features of a time series:

1) overall, if $d<0.5$, $X_t$ is stationary.

2) if $d\in(-1,-0.5)$, the series $X_t$ exhibits invertibility.

3) if $d\in(-0.5,0)$, the stationary $X_t$ is antipersistent.

4) if $d=0$, the stationary $X_t$ are short memory and meanreverting process.

5) if $d\in(0,0.5)$, $X_t$ exhibits long-memory positive dependence.

6) if $d\in(0.5,1)$, $X_t$ is mean reverting but may not be stationary.

7) when $d=1$, $X_t$ is a unit root process.

The hybrid model (ARFIMA-GARCH) can explain heteroscedasticitybetween different observations. Restrictions on constant variance in ARFIMA are relaxed by replacing $\varepsilon_t$ with the following $\xi_t$.

$$\xi_t=\sigma_t\varepsilon_t \tag{3}$$

$$\sigma_t^2=\alpha_0+\sum_{i=1}^{m}\alpha_i\xi_{t-i}^2+\sum_{j=1}^{n}\beta_j\sigma_{t-j}^2 \tag{4}$$

where n is the order of $\sigma_t$, called the GARCH term, and m is the order of $\xi_t$, called the ARCH term; $\xi_t$ represents the error term; $\varepsilon_t$ denotes the independent distribution random term with zero mean and unit variance. This paper chooses the exponential GARCH ( EGARCH ) model, then the above equation can be written as:

$$\ln\sigma_t^2=\left(\alpha_0+\sum_{k=1}^{l}\zeta_k v_{kt}\right)+\sum_{i=1}^{m}\left(\alpha_i z_{t-i}+\gamma_i\left[|z_{t-i}|-E|z_{t-i}|\right]\right)+\sum_{j=1}^{n}\beta_j\ln\sigma_{t-j}^2 \tag{5}$$

## 4. The Empirical Results

### 4.1. Descriptive Statistics

The time series data used in this study is the monthly unemployment rate in Hong Kong from October 1982 to August 2022. The basic descriptive statistics are shown in Table 2. A total of 479 observations (months) were obtained. The monthly average unemployment rate is 3.764%, the minimum monthly unemployment rate is 1%, and the maximum monthly unemployment rate is 8.5%, which is quite different.

**Table 2.** Descriptive statistics of unemployment rate

| | | | | |
|---|---|---|---|---|
| N | 479 | Mode | | 3.300 |
| Mean | 3.764 | Skew | Skewness | 0.606 |
| Std. Dev. | 1.689 | | Std. Error of Skewness | 0.112 |
| Min | 1.000 | Kurt | Kurtosis | -0.200 |
| Max | 8.500 | | Std. Error of Kurtosis | 0.223 |
| Median | 3.400 | | | |

## 4.2. ARFIMA-EGARCH Model Estimation

Perform ADF and KPSS tests on the unemployment rate sequence, and the results are shown in Table 3. Both ADF and KPSS test results indicate that the original sequence is non-stationary.

**Table 3.** Unit Root Test

| | P-value |
|---|---|
| ADF | 0.23 |
| KPSS_Level | 0.01 |
| KPSS_Trend | 0.01 |

The p-value of the pure randomness test on the original sequence is equal to 0.000, therefore rejecting the original hypothesis indicates that the original sequence is not random. Using the ARFIMA(1,d,1) model to fit the original sequence, we obtained $d = 0.445$, indicating that the original sequence is a fractional integration process with strong long-term memory. In order to determine whether there is conditional heteroscedasticity in the perturbed terms of the original sequence, we conducted LM and PQ tests on them. The p-value of the LM and PQ tests with a 5-order lag is equal to 0, therefore rejecting the original hypothesis and believing that the original sequence has an ARCH effect.

The original sequence is a fractional integration process with long memory. In order to capture and adapt to long memory characteristics, ARFIMA was selected as the mean model and EGARCH as the variance model. For simplicity of calculation, except for the integer order of the difference, the order of each component in the model is set to 0 or 1. In addition, in order to select the most suitable model structure to fit data and explore the characteristics of time series, seven model variants, ARFIMA(1,d,1)- EGARCH(1,1), ARFIMA(0,d,1)-EGARCH(1,1), ARFIMA(1,d,0)-EGARCH(1,1), ARFIMA(1, d,1)-EGARCH(0,1), ARFIMA(0,d,1)-EGARCH(0,1), ARFIMA(1,d,0)-EGARCH(0,1) and ARMA(1,1), were developed.

In order to verify whether the original sequence is a normal distribution, we conducted the Kolmogorov Smirnov test, and the result rejected the original hypothesis that the unemployment rate sequence belongs to a normal distribution. Therefore, in the following model, the error term is assumed to follow a skewed normal distribution.

**Table 4.** Performance standards of the model

| Models | AIC | BIC | SIC | HQIC | Loglikelihood |
|---|---|---|---|---|---|
| ARFIMA(1,0.428,1)-GARCH(1,1) | -1.1381 | -1.0597 | -1.1388 | -1.1073 | 281.5699 |
| ARFIMA(0,0.500,1)-GARCH(1,1) | -0.4476 | -0.3779 | -0.4481 | -0.4202 | 115.2045 |
| ARFIMA(1,0.396,0)-GARCH(1,1) | -1.1407 | -1.0711 | -1.1413 | -1.1133 | 281.2069 |
| ARFIMA(1,0.500,1)-GARCH(0,1) | -0.8823 | -0.8214 | -0.8827 | -0.8584 | 218.3157 |
| ARFIMA(0,0.500,1)-GARCH(0,1) | -0.0660 | -0.0138 | -0.0663 | -0.0455 | 21.8162 |
| ARFIMA(1,0.500,0)-GARCH(0,1) | -0.8803 | -0.8281 | -0.8806 | -0.8598 | 216.8393 |
| ARMA(1,1) | -361.2264 | -348.7113 | -361.2012 | | 183.61 |

The performance standards of 7 different models are compared in Table 4. The results show that the ARFIMA(1,0.428,1)-EGARCH(1,1) and ARFIMA(1,0.396,0)- EGARCH(1,1) models outperform other models due to their smaller information criterion values and larger LL.

Perform weighted Ljung Box and ARCH-LM (Lagrange Multiplier) tests on the ARFIMA(1,d,0)-EGARCH(1,1) and EGARCH(1,1) models, respectively, to test the efficiency of the proposed methods in considering overall time correlation and heteroscedasticity. The results are shown in Table 5. The statistics of the Ljung Box test reject the original hypothesis, indicating that the square residuals in the EGARCH(1,1) sequence are not independently distributed. However, ARFIMA(1,0.396,0)-EGARCH(1,1) does not have sequence related squared residuals. The ARCH-LM test results with a lag of 5 periods cannot reject the original hypothesis, indicating that the ARFIMA(1,0.396,0)-EGARCH(1,1) model does not have the ARCH effect.

**Table 5.** Weighted Ljung Box and ARCH-LM tests

| | ARFIMA(1,d,0)-EGARCH(1,1) | EGARCH(1,1) |
|---|---|---|
| Models | d=0.396 | |
| Weighted Ljung-Box Test on Standardized Squared Residuals | 1.4761 | 253.59*** |
| Null: No serial correlation | P-value=0.7458 | P-value=0.000 |
| Weighted ARCH LM Tests | 1.1159 | 1.667*** |
| Null: No Arch effect | P-value=0.6989 | P-value=0.000 |

*** at 1% level of significance.

## 5. Conclusion

The prediction of unemployment rate is based on historical unemployment rate data. Capturing long-term memory correlations in historical unemployment rate time series is crucial for accurately predicting future unemployment rates. However, this long-term memory characteristic is often overlooked in literature. To fill this research gap, an autoregressive score integral moving average model with generalized autoregressive conditional heteroscedasticity (ARFIMA-GARCH) was used to capture and adapt to neglected long-term correlations in unemployment rate data.

This study is based on monthly unemployment rate data from Hong Kong for the past 40 years, and estimates a total of 7 ARFIMA-EGARCH models for comparison. The results showed that the difference order ( $d = 0.396$ ) of the best specified model ARFIMA(1,d,0)-EGARCH(1,1) indicates the existence of long memory in the time series of unemployment rates. The current high unemployment rate will also have a significant impact on future employment. Life does not have long-term stability, let alone China is in a new historical period of achieving the first centenary goal and moving towards the second centenary goal. Employment forms will become more severe due to various pressures, leading to an increase in unemployment rate. Therefore, in order to avoid the risk of more severe unemployment in the future, we must adhere to the instructions and requirements of the Party Central Committee and strive to reduce the unemployment rate.

## References

[1] F. Cai, Y. Du, W. Gao: Employment elasticity, natural unemployment, and macroeconomic policies-why hasn't economic growth brought explicit employment? Economic Research, Vol. 9 (2004) No.47, p.18-25.

[2] X.Q. Zeng, Y. Yu: Measurement and analysis of China's natural unemployment rate, Chinese Social Sciences, Vol. 4 (2006) No.206, p.65-76.

[3] C.H. Su, Y. Zhang, S. Li: The impact of delayed retirement age on China's unemployment rate: theory and verification, Journal of Shandong University (Philosophy and Social Sciences Edition), Vol. 1 (2015), 11-22.

[4] Z.Z. Han, N.N. An: Minimum wage, labor supply, and unemployment: an empirical analysis based on the VAR model, Journal of Jinan (Philosophy and Social Sciences Edition), Vol. 1 (2007), 38-44.

[5] Y.F. Zhu, R. Liu, J. Ye: The impact of economic new normal and trade protection on unemployment in various countries: an empirical study based on the GVAR model, Business Economics Research, Vol.3 (2019), 133-135.

[6] Y. Zhang, S.Z. Jiang: Analysis of the interaction between unemployment insurance and employment status: based on the SVAR model, China Insurance, Vol. 6 (2018), 41-46.

[7] J.P. Li: Empirical study on the relationship between economic growth, foreign trade, and urban unemployment in China: based on cointegration model and variance decomposition, Journal of Shanxi Finance and Taxation College, Vol. 12 (2010) No. 3, p.67-70.

[8] L. Yang: Research on the relationship between urban rural labor flow and urban unemployment based on VAR model, Labor Economy Review, Vol.6 (2013) No. 0, p.72-80.

[9] B. Zhang: Analysis of the impact and countermeasures of urban-rural income gap on urban unemployment, Urban Issues, Vol. 2 (2007) No. 71, p.45-50.

[10] Y. Ding: Dynamic research on the relationship between rmployment and resident income distribution (Ph.D., Jinan University, China, 2008), p.219.

[11] D.L. Xu, R.H. Wang: Comparative analysis of registered unemployment and surveyed unemployment, Statistics and Decision, Vol. 9 (2009), 14-16.

[12] G.Y. Wang: Several predictive models for urban registered unemployment rates in Shandong Province (MS., Shandong University, China, 2021), p.59.

[13] Z.W. Yuan: Analysis and Prediction of Factors Influencing China's Real Unemployment Rate (MS., Liaoning University, China, 2021), p.52.

[14] Y. H. Li, Y. Wei, B.Z. Zhang: Research on the long memory characteristics of historical information in the stock market, China Management Science, Vol. 23 (2015) No. 9, p.37-45.

[15] J.L. Liu, J. Qu: Research on long memory of China's commodity futures price returns and volatility, Business Research, Vol. 1 (2017), 57-68.