# Forecasting Foreign Trade Trends Based on Combined ARIMA and Composite Quantile Regression Models

Luwei Zhang

Zhanglw35@mail2.sysu.edu.cn

School of Mathematics (Zhuhai), Sun Yat-sen University, Zhuhai, Guangdong, China

**Abstract.** Predicting the trend of total foreign trade is vital for studying the national economic system. In this paper, after using a full subset regression model for variable screening, a combined model of ARIMA and composite quantile regression (CQR) estimation series is used to predict the trend of total foreign trade. This model combines internal and external factors to make forecasts, which has higher accuracy and stronger stability over a short period at specific accuracy values, and can obtain better forecasting results than a single model. In practice, the combined model used in this paper can give full play to the influence of specific factors, thus providing a more accurate and robust forecasting model for scientific and rational decision-making.

**Keywords:** ARIMA model, composite quantile regression, foreign trade, trend forecasting

## 1.Introduction

Total foreign trade, as a macroeconomic indicator, plays an important role in the study of national economic systems and foreign economic trade trends. In the study of forecasting financial time series, common single analysis methods have their advantages and disadvantages, and Bates and Granger (1969) proposed that the combined forecasting method of using different forecasting methods on the same time series and averaging the forecasting results tends to improve the forecasting accuracy [1]. Zhang and Sun (2019) used ARIMA and SVM combined models to analyze China's total imports and exports and obtained through empirical analysis that the combined model can achieve more accurate estimation in total import and export forecasting compared to a single model [2]. The autoregressive integrated moving average model (ARIMA) uses the historical data of the variables themselves for direct forecasting, while the regression analysis can take into account the effects of multiple explanatory variables, and it has some theoretical significance and practical value to combine the two and apply them in the forecasting of time series.

To circumvent the estimation drawbacks of the least squares method, Koenker and Bassett (1978) proposed a more robust quantile regression estimation method, which can portray the complete conditional distribution through the selection of different quartiles [3]. Based on this, Zou and Yuan (2008) first defined the composite quantile regression method, which can gather more valid information by considering multiple quantiles simultaneously and tends to exhibit better estimation properties [4]. Zhao and Chen (2016) combined the composite quantile regression estimation method with a GARCH model to build an estimation model with stock prices as the study object and obtained significantly better results than previous estimation methods [5].

MSEA 2023, May 26-28, Nanjing, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.26-5-2023.2334441

In this paper, we use a full-subset regression model to filter out the set of independent variables, and then use a composite quantile regression estimation series method to forecast, using a combined forecasting method that averages the forecast values of ARIMA and regression models, which can obtain a better forecasting effect than a single model. In addition, the model we use combines internal and external factors to make forecasts, which can give full play to the influence of specific factors in practical applications and achieve more effective estimation of the trend of the dependent variable.

## 2.Methods

### 2.1.ARIMA-based time series forecasting model

#### 2.1.1. Selection of models

The ARIMA model is popular in the fields of economy and finance. Studies have shown that the model is suitable for short-term forecasting, and it is often more effective for forecasting the next two years for a data set in years. In this paper, we use 42 annual data of China's total foreign trade from 1978-2019 to form a time series to forecast the trend of the following two years.

The general form of the ARIMA(p,d,q) model for a smooth time series is

$$x_t = \beta_0 + \beta_1 x_{t-1} + \cdots + \beta_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \ . \tag{1}$$

where $x_t$ is the observed value of the time series at time t, p and q are the number of autoregressive terms and moving average terms of the time series, respectively, d is the number of differences needed to transform the time series into a smooth time series,$\{\beta_t\}$ and$\{\theta_t\}$ parameters, which are a white noise series.

#### 2.1.2.Stability check

Only a smooth time series can fit the ARIMA model, so it is necessary to differentially process the non-smooth time series and test for smoothness before modeling. Determining smoothness based on the autocorrelation plot of the time series is usually highly subjective, so we use the unit root test (ADF test) in this paper.

For the first-order difference of the time series after the ADF test to obtain the series is non-stationary, followed by the second difference and ADF test, the p-value is about 0.01, less than 0.05, indicating that the second difference of the series is a stationary series. This paper also uses the "ndiffs" function of R software to determine that at least the second-order difference is a smooth series.

#### 2.1.3.ARIMA model development and testing

ARIMA(0,2,1) is the selected fitted model, and the order of its difference is under the above discussion. The significance test of the selected ARIMA model is to determine whether the residuals are white noise. If they are white noise series, then the residuals are partly pure random series that cannot be captured so that the model fits well.

Figure 1 shows that the autocorrelation coefficients are all within the bounds. The p-value is

0.6111 from the LB test, which is significantly greater than 0.05, indicating that the residual series is purely random. The ARIMA(0,2,1) model passes the significance test.

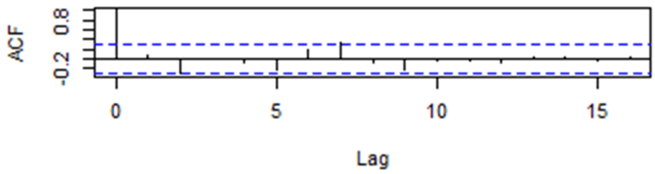


**Figure 1.** Autocorrelation test for residual series (ACF test)

### 2.2.Time series forecasting model based on (composite) quantile regression

#### 2.2.1.Variable Selection

By reviewing a large amount of literature, we selected ten factors that affect the trend of total foreign trade: gross national income (X1), gross domestic product (X2), population consumption level (X3), total population at the end of the year (X4), the actual amount of foreign capital utilized (X5), the exchange rate of RMB to USD (X6), tariff (X7), fiscal revenue (X8), fiscal expenditure (X9), foreign exchange reserves ( X10). Data from the National Bureau of Statistics of the People's Republic of China.

There are two common methods to select the final valid set of independent variables from multiple factors: the stepwise regression method and the full subset regression method. The stepwise regression method evaluates a limited number of models and does not guarantee that the model obtained is the best. However, the full-set regression method captures the interaction effects between the independent variables while comparing all possible models.

To combine with the ARIMA model, we also use the data from the first 42 years to estimate the total foreign trade for 2020 and 2021. After treating the missing data values in the independent variables using linear interpolation, we run the functions in the R software leaps package and the MASS package to perform full subset regressions and stepwise regressions for the data sets of the ten independent variables and the dependent variable total foreign trade for X1 to X10, respectively, from 1978 to 2019. Both methods screened X2, X3, X4, X5, X7, X8, and X9 as the best set of independent variables, where the screening results of the full subset regression method are shown in Figure 2.

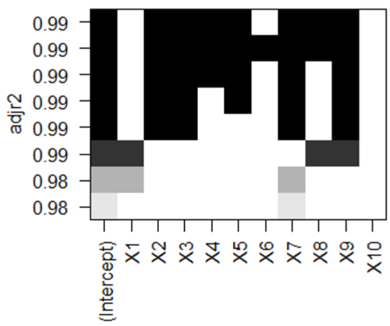


**Figure 2.** Full subset regression results

When using the regression model to do the forecast estimation, we approximate the exact values of the independent variables in 2020 and 2021 as their forecast values for ease of calculation. After that, we use the regression model to estimate the forecast values of total foreign trade.

### 2.2.2.Quantile regression estimation

The estimated equation for the quantile regression is

$$\left(\hat{b}_\tau, \hat{\boldsymbol{\beta}}^{QR_\tau}\right) = \underset{b,\boldsymbol{\beta}}{\arg\min} \sum_{i=1}^{n} \rho_\tau\left(y_i - b - \sum_{j=1}^{p} x_{ij}\beta_j\right). \tag{2}$$

where $0 < \tau < 1$, $\rho_\tau(u)$ are loss functions.

### 2.2.3.Compound quantile regression estimation

The estimated equation for the composite quantile regression is

$$\left(\hat{b}_1, \cdots, \hat{b}_K, \hat{\boldsymbol{\beta}}^{CQR}\right) = \underset{b_1,\cdots,b_k,\boldsymbol{\beta}}{\arg\min} \sum_{k=1}^{K} \left\{ \sum_{i=1}^{n} \rho_{\tau_k}(y_{\mathrm{i}} - b_k - \mathbf{x_i}^T \boldsymbol{\beta}) \right\}. \tag{3}$$

where K is the precision and the quantile $\tau_k = \frac{k}{K+1}$ for $k = 1, 2, \cdots, K$, $\rho_{\tau_k}(u)$ is the loss function.

## 2.3.Combined forecasting model of ARIMA and (composite) quantile regression

### 2.3.1.The selection of K

Zou and Yuan (2008) concluded that the estimation is better for the accuracy value K 9 and K=19 is a good choice [4], Xu et al. (2017) selected K=5,9,19 to study the combined model of composite quantile regression and neural network [6], according to previous studies, the best selection range of accuracy value K is concentrated between 5 and 19. In this paper, we select the accuracy values K=5,9,11,19 to study.

### 2.3.2.Establishment of the optimal model

Pietrosanu and Gao (2021) compared quantile and composite quantile regression models based on four algorithms: the alternating direction method of multipliers (ADMM), majorize-minimization (MM) coordinate descent (CD), and interior point (IP) algorithms, and demonstrated that the MM method is best suited for non-regularized (composite) quantile regression, while in regularized quantile regression, all method have similar performance in terms of variable selection [7]. Therefore, we use the "rq" function from the "quantreg" package in the R software to fit the quantile regression (QR) model and the "cqr.lasso.mm" function from the "cqrReg" package to fit the composite quantile regression (CQR) model. We then take the average of the predicted values of the ARIMA model and the above two models, respectively, as the predicted values of the combined model.

For a given accuracy K and a specific forecast year, we use the model to calculate the estimated values of the combined model at different quartiles and select the one with the smallest absolute error from the exact value as the forecast value of the optimal model.

## 3.Results

### 3.1.Prediction results of ARIMA model

We used the ARIMA(0,2,1) model that passes the smoothness test and pure randomness test in R software to fit the total foreign trade time series, and the ARIMA model prediction trend graph is shown in Figure 3.

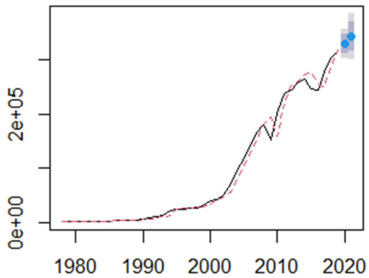


**Figure 3.** ARIMA(0,2,1) prediction image

Table 1 shows the forecast values and error results for 2020 and 2021.

**Table 1.** Comparison of prediction results of ARIMA model

| year | Accurate value | ARIMA | ARIMA error |
|---|---|---|---|
| 2020 | 322215.2 | 328969.8 | 6754.6 |
| 2021 | 390921.67 | 342312.3 | -48609.37 |

Note: ARIMA error is the difference obtained from the ARIMA(0,2,1) prediction minus the exact value of the corresponding year

### 3.2.Prediction results of combined ARIMA and (composite) quantile regression models

In the four cases with precision values K=5,9,11,19, we used quantile regression (QR) and compound quantile regression (CQR) models to predict the dependent variable (total foreign trade). After that, we averaged the predicted values of the ARIMA model and the predicted values of the above two models, respectively, as the predicted values of the combined model.

As shown in Table 2, the combined model provides a more accurate prediction than the ARIMA single model. The combined model with ARIMA and CQR at K=11 has the slightest mean absolute error, and the prediction effect is the most accurate overall. The combination of ARIMA and CQR has more powerful stability than its combination with QR.

**Table 2.** Comparison of prediction results of ARIMA model

| | K=5 | | K=9 | | K=11 | | K=19 | |
|---|---|---|---|---|---|---|---|---|
| | QR | CQR | QR | CQR | QR | CQR | QR | CQR |
| 2020 optimal quantile | 2\6 | 5\6 | 4\10 | 8\10 | 4\12 | 10\12 | 8\20 | 17\20 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2020 combination error | -1362.2 | -3194.140205 | -1150 | -408.5708981 | -1362.2 | 82.03681193 | -1150 | -887.3420104 |
| 2021 optimal quantile | 1\6 | 5\6 | 2\10 | 8\10 | 2\12 | 7\12 | 4\20 | 17\20 |
| 2021 combination error | 10.23 | -107.3396109 | 10.23 | -34.70478474 | 10.23 | 73.21754321 | 10.23 | -960.9969859 |
| Mean absolute error | 686.215 | 1650.739908 | 580.115 | 221.6378414 | 686.215 | 77.62717757 | 580.115 | 924.1694982 |

Note: K is the precision value and the mean absolute error is the two-year average of the absolute error under the corresponding combined model. The optimal quantile is the quantile that minimizes the absolute error between the predicted and actual values of the combined model.

## 4.Discussion and conclusion

This study shows that the combined ARIMA(0,2,1) and composite quantile regression model with accuracy value K=11 has high accuracy and strong stability in forecasting total foreign trade over a short period. The combined model can combine the advantages of the two different types of forecasting methods and thus reduce the forecasting error. ARIMA performs well for short-term forecasting while the composite quantile regression method is more robust.

This combined model combines internal and external factors to make forecasts, which can tap into the trend of the dependent variable itself and also reasonably incorporate the influence of the explanatory variables into the forecast analysis, thus providing a reasonable explanation of the forecast trend based on the relationship between the independent and dependent variables.

The application of the prediction method of this combined model in other fields needs to be further investigated, and the value of K in the (composite) quantile regression method needs more attempts. In further research, we will improve and optimize the model proposed in this paper, and combine more models as weighted composite quantile regression with ARIMA to make more accurate and robust forecasts. Besides, we will apply the combined model to other fields of time series forecasting to verify the generalization value of the model.

## References

[1] Bates, J. M., & Granger, C. W. J.: The combination of forecasts. Operational Research Quarterly, 20(4), 451–468 (1969)

[2] Zhang Lei, Sun Deshan.: Forecasting total import and export trade based on portfolio model[J]. Jiangsu Business Theory,2019(02):57-59+66.DOI:10.13395/j.cnki.issn.1009- 0061.2019.02.011.

[3] Koenker, R. , & Bassett, G. W.: Regression quantiles. Econometrica, 46 (1), 33–50 (1978)

[4] Zou, H. , & Yuan, M.: Composite quantile regression and the oracle model selection theory. Annals of Statistics, 36 (3), 1108–1126 (2008)

[5] Zhao B, Chen Z, Tao G, etc.: Composite quantile regression estimation for P-GARCH processes[J].Science China(Mathematics), 59(05):977-998 (2016)

[6] Xu, Q., Deng, K., Jiang, C., Sun, F., & Huang, X.: Composite quantile regression neural network with applications. Expert Systems with Applications, 76, 129–139 (2017)

[7] Pietrosanu, M., Gao, J., Kong, L. et al.: Advanced algorithms for penalized quantile and composite quantile regression. Comput Stat 36, 333–346 (2021)